

Bryan Levman*

The language of early Buddhism

DOI 10.1515/jsall-2016-0001

Abstract: The earliest recoverable language of Buddhism is a Middle Indo-Aryan (MIA) *koine*, an inter-dialect language which reduced linguistic variability by dialect levelling and simplification, through elimination of interdialect phonological differences which impede understanding, and harmonization of the different dialects to a common language intelligible across all dialects. Among the principal features of this *koine* the following can be identified: reduction of the distinction between voiced and unvoiced intervocalic stops; replacement of aspirated stops by aspirates only; replacement of diphthongs by a simple vowel; the assimilation or resolution of most consonant clusters; the merging of dental, retroflex and palatal sibilants; the interchange of a variety of liquids and labials. Among the various factors which led to the *koine*'s formation, were other Indo-Aryan and non Indo-Aryan languages and *koine* whose phonetic constraints catalyzed and accelerated the development of MIA in certain directions, that is towards harmonization with their own phonological system which lacked such features as phonemic intervocalic voiced stops, aspirated stops, consonant clusters, etc., and had other features (like retroflexes) which Indo-Aryan lacked.

Keywords: early Language of Buddhism, *koine*, *lingua franca*, linguistic diffusion

1 What language(s) did the Buddha speak?

One of the big mysteries of Buddhism is the language the Buddha spoke and what his actual words were. This is a question which has preoccupied Buddhist scholars for centuries. As is well known, Buddhaghosa equated Pāli (P) with Māgadhī,¹ but we know that Pāli is a composite dialect, and although it contains elements of what is probably an eastern dialect that the Buddha may have

¹ *Samantapāsādikā* (commentary on the *Vinaya*), 1214^{18–19}: *ettha sakā nirutti nāma sammāsambuddhena vutta-ppakāro Māgadhika-vohāro*, translated by Horner (2001 [1938]: 194, footnote 1), as “the current Magadhese manner of speech according to the awakened one,” but see Levman (2008–2009: 35), for discussion on the meaning of *nirutti*. When referring to Buddhist canonical works, the page and line numbers (in superscript) refer to the Pali Text Society editions.

*Corresponding author: Bryan Levman, Visiting Scholar, University of Toronto, Department for the Study of Religion, E-mail: bryan.levman@utoronto.ca

spoken, it is nevertheless not an “original language of Buddhism” but a translation of something earlier. It is usually characterized as a western dialect, but in fact, if closely analysed, it contains elements of both eastern, western, and northwestern dialects – it is a mixed language created by monks, normalized for religious purposes (Norman 1983: 4; von Hinüber 1983: 1–9; Lamotte 1988: 563; Levman 2014: 43–46).

Most scholars have assumed that the Buddha imparted his teachings in either Māgadhi or Old Ardhamāgadhi (Lüders 1954: 7–8), Old Māgadhi (Norman 1980b: 71),² or Ardhamāgadhi (AMg, Alsdorf 1980: 17–23). However, this is probably too narrow a view. Certainly, the Buddha had close ties with the Magadha kingdom and much of his ministry was spent there. However, neither his home town (Kapilavatthu) nor his death place (P Kusinārā, Sanskrit [S] Kuśinagarī), nor Bārāṇasī (S Vārāṇasī), Vesālī (S Vaiśālī) nor Sāvatti (S Śrāvastī) where he delivered many of his sermons are in Magadha (Edgerton 1988[1953]: 3, footnote 8; Roth 1980: 78). Bārāṇasī was the former capital of the Kasi kingdom (eventually assimilated by Ajātasattu, king of Magadha), and Sāvatti was the capital of the Kosalan kingdom of which the Buddha’s Sakyan clan were vassals. Rhys Davids says that the Buddha spoke “Kosalan” (1908: 3), but of course we don’t know what that is. But the other locations were all tribal capitals – Kapilavatthu (S Kapilavastu) was the capital of the Sakyan tribe, Kusinārā, the Mallas and Vesālī, the capital of the Licchavi tribal federation. Although we know little about the languages of the clans, we can be fairly certain that they spoke a non-Indo-Aryan language because most of the place names in the *gaṇasaṅghas* – the republics of the clans – are non IA in origin and many of the botanical names and words associated with cultural practices (like funerals) are autochthonous (Levman 2013: 148–149). Whether the Buddha spoke a non IA language or not, we do not know; nevertheless there is lots of linguistic evidence that the phonology of the pre-existent languages affected the Middle Indic dialects of Pāli and the Prakrits (see below).

Regardless of which language or languages the Buddha taught, we know that very early on, perhaps in his lifetime, or shortly thereafter, as his teachings spread rapidly across India, they became less and less understandable in their original form to speakers of other dialects. We know, for example, that the north-western dialect (Gāndhārī [G]) was quite different from the eastern dialects, and it was unlikely that they were mutually comprehensible. Certainly by Aśoka’s time the dialect differences between, for example, the dialect of Shābāzgarhī in the north and Kālsī in the east were considerable. By comparing parallel Buddhist translations that have come down to us in the different dialects, we can identify an

² Norman believed that some of the Buddha’s teachings must have been in Old Māgadhi, but that “there was no single language or dialect used by the Buddha for his preaching” (1980b: 75).

underlying earlier layer which one might characterize as a common language or *koine*, an inter-dialect or interlect form with all the principal dialect differences removed and homogenized for ease of communication across dialect boundaries. The nature of this *koine* I have discussed in detail elsewhere (Levman 2014: 460–465); It is characterized, *inter alia*, by lenition or elimination of intervocalic stops, lenition of aspirated stops to aspirates only, assimilation or resolution of consonant clusters, levelling of sibilants, and interchange of glides, nasals, palatals and liquids; it is most similar to the G dialect, but not identical.

2 Historical scholarship

Sylvain Lévi (1912) was the first to note the existence of *une langue précanonique*; he identified an earlier layer underlying P and the Sanskritized Prakrits which were the “inheritors of an earlier tradition, recited or compiled in a dialect which has disappeared and which had already attained an advanced stage of phonetic change” (511). Geiger (1916: 3–4) called it a *lingua franca* which contained elements of all the dialects, but was free from the most obtrusive dialectal characteristics. He described it as a language of the higher and cultured classes (*Hoch und Gebildetensprache*) which had been brought into being already in pre-Buddhistic times through the needs of inter-communication produced by social interaction (a *Verkehrssprache*). Geiger believed that P was a form of Māgadhi or AMg as actually used by the Buddha – not a pure Māgadhi, but one which avoided the grossest dialectal peculiarities of the language; that is a new, artificial language which evolved out of the language of the teacher. Edgerton (in 1936) called it “an earlier dialect, Prakritic in character, in which there must have existed at one time a considerable body of (perhaps only oral) Buddhist literature (502).” Both P and Buddhist Hybrid Sanskrit (BHS or Sanskritized Prakrit) were translations from it. In 1952 Helmer Smith identified this early language of Buddhism as a *koine gangétique* of which Pāli and Ardha Māgadhi represented the oldest normalizations (“...dont l’ardhamāgadhi et le pali représentent les normalisations les plus anciennes...” p. 178). Lüders believed that an “Urkanon” lay at the base of the Pāli and Buddhist Sanskrit writings composed in an eastern dialect but further evolved; he identified the phonology of the Urkanon with Aśoka’s *Kanzleisprache* (the administrative language of the ruling government in Pāṭaliputra),³ but at a higher degree of development (“auf einer weiteren Stufe der

³ Prof. Max Deeg has also helpfully pointed out the origin of this (anachronistic) term in the early middle ages of Germany where a *Kanzleisprache* was the official language of the court which suppressed dialect differences in order to communicate over a wide geographical area.

Entwicklung”, 1954: 8) where voiceless stops were softened and voiced stops disappeared, to name two of the characteristic features of the underlying language. In 1983 von Hinüber called the underlying language Buddhist Middle Indic, a common Buddhist language from which both P and Buddhist Sanskrit branched off, but one which he believed was later than the earliest language of the canon (1983: 192–193). Whether this was a *lingua franca* or *koine* as Geiger and Smith opined or an earlier lost dialect or sub-dialect as Norman has suggested (2006: 95) is impossible to tell; however, the malleability and flexibility of the language suggests to me that it indeed was a *koine* which must have existed at that time in north India for trade and administrative purposes.

3 Definitions

In this paper I use *lingua franca* in the generic sense of “any language that is used by speakers of different languages as a common medium of communication; a common language” (OED). As is well known, the “original” *lingua franca* was a trading inter-language of the Mediterranean utilizing a mixture of Italian, Provençal, Spanish, Arabic, Greek and Turkish in the late middle ages. The word *koine* (from Gk. *κοινή*, ‘common’) is often used as a synonym for *lingua franca*, but here I use it in the more technical sense of an inter-dialect language that comprises features of several regional varieties, but is based primarily on one of them, in a reduced and simplified form. “Functionally, the original [Greek] *koine* was a regional *lingua franca* which became a regional standard. It was spoken mostly as a second language but did become the first language of some” (Siegel 1985: 358–359). A *koine* is a new dialect which reduces linguistic variability by reduction of the forms available; this reduction takes place through “a process of koinéization, which consists of the *levelling* out of minority and otherwise marked speech forms, and of *simplification*, which involves, crucially, a reduction in irregularities” (Trudgill 1986: 107; italics in original). This definition perfectly describes the linguistic form underlying the Buddhist translations – in P, Gāndhārī (G), and the other Prakrits partially or completely Sanskritized – that have come down to us. The term “marked” describes features that are in a minority in the mix in terms of the number of speakers who use them, or have a restricted regional currency (Trudgill 1986: 98).

4 Methodology

The features and nature of the common language may be isolated by standard comparative linguistic techniques; that is, by comparing cognate forms and

reconstructing what the underlying form must have been, based on the principles of directionality, shared features and economy (Campbell 2004: Ch 5, 122–167). Lévi, for example, identifies the weakening of intervocalic consonants as one of the principal phonetic characteristics of this “langue précanonique du bouddhisme”. This led to misunderstandings and wrong translations (hyperforms) when the word was translated back into S or a Sanskritized Prakrit like P. These hyperforms provide the best evidence for demonstrating that P was the result of a translation process from a more simplified, earlier stratum whose meaning has sometimes been lost (Norman 1989: 375). On a step-by-step basis this involves (1) comparing corresponding passages in the different transmissions of the teachings (Prakrit, Sanskritized Prakrit and S), (2) isolating words that are phonologically cognate and noting differences, (3) extrapolating the underlying form that would logically give rise to the differences in these forms, based on normal historical linguistic sound changes, and (4) explaining the anomalies in terms of hyperforms, that is, misinterpretations of the meaning of a source word, or simply variant translations (Sanskritizations) of an underlying word that is ambiguous in meaning because of the dialect levelling that has taken place.⁴ See below for examples. None of this proves *per se* that the underlying form is koinic; it could just as well be an earlier, now lost MI dialect. However, when one looks at all the other evidence assembled here, especially the structure of *koinés*, the need for some kind of inter-language in mid first millennium ancient India, the existence and similarity of other non IA common languages, etc., the *koiné* hypothesis is compelling.

4.1 Examples

Take, for example, the compound describing the famous park where the Buddha preached his first sermon, which Lévi notes as an example of intervocalic consonant weakening (1912: 499–500); the name has been preserved in two forms, as *isi-patana* (‘descent of the seers’) and *isi-vadana* (‘speaking of the seers’); however, as Caillat has shown (1968: 177–183), the underlying form which gave rise to both these possibilities was *isi-vayana* which itself was derived from S *ṛṣya-vṛjana* (‘antelope enclosure or pasture’), where *ṛṣya* > *isi* and *vṛjana* > *vajana* > *vayana*. The translators had misconstrued *isi* as derived from S *ṛṣi* (‘seer’) and *vayana* as derived from S *patana* (‘descending; the

⁴ Norman (1989: 375) defines hyperforms as “... forms which are unlikely to have had a genuine existence in any dialect, but which arose as a result of bad or misunderstood translation techniques”. For a fuller discussion on methodology see Levman (2014: Chapter 3, 79–108).

weakening of *-p-* > *-v-* being a common Old Indic [OI] > Middle Indic [MI] change), or *vadana* ('talking'). Interestingly, the commentary retains the correct etymology of the compound (*migānaṃ abhaya-dāna-vasena*, 'because of being a fearless retreat for animals') while inventing fake etymologies for *isi-patana* and *isi-vadana* (Levman 2014: 394–396).

The use of the intervocalic *-y-* glide as a substitute for a voiced consonant was quite common in the underlying language. This feature was preserved in the AMg language of the Jains, but (in most cases) back-translated in the other traditions. In one of the earliest of the Buddhist transmissions, the *Sutta Nipāta* (Sn), parts of which may actually go back to the historical Buddha, we find the word *virajo* ('free from corruption') in the Pāli version and *virato* ('ceased') in the BHS version of the poem (*Mahāvastu*). The underlying form must have been transmitted as *virayo*, where the intervocalic consonant was replaced with the *-y-* glide, leaving each translator to insert what he/she opined to be the "correct" consonant (Norman 1980a: 148–161; Levman 2014: 245–257).

These phonetic simplifications are exactly what we would expect to find in the natural evolution of OI > MI, and the further levelling of MI inter-dialect differences to make them more readily mutually comprehensible. Some MI dialects tended to voice intervocalic stops (e.g. G), some to devoice them (Paiśācī), and others to eliminate them altogether (AMg). An inter-language common to all three tendencies would eliminate the stop or replace it with a glide; this would facilitate not only inter-dialect communication, but also inter-language intelligibility with other speakers whose language lacked the voiced-voiceless phonemic distinction (see below). The weakening or elimination of intervocalic stops was equally applicable to aspirated stops, where the same proclivity for confusion arose, both between MI dialects where pronunciations differed and other languages which lacked the aspirated stop phoneme.

A common epithet of the Buddha like *tathāgata* (usually translated 'thus come' = *tathā-āgata* or 'thus gone' = *tathā-gata*), for example, appears in AMg as *tahāgaya*, where the aspirated stop > aspirate only and the intervocalic stop > glide. This was probably also close to the *koine* form, as the characteristic dialect differences have been removed and it is up to the hearer to replace the aspirate and glide with the relevant phoneme from his/her dialect. So whether one pronounced it *tathāgata*, *tathāgada*, *tathāgasa*, *tahāgaya*, etc., in one's own dialect, the simplified *koine* form (*tahāgaya* or *tahāga'a*; ' = > Ø) would be a common denominator of all forms, that is the most reduced form from which all can be derived. The latter part of this compound *-gaya* could of course point to another meaning altogether (with the intervocalic *-y-* taken as a glide phoneme, substituted for a weakened stop), as Buddhaghosa theorizes in his commentary, i.e. *-āgado*, 'speech' (*tatha-āgada*, 'true speech'), or *-agada*, (*tatha-agada*, 'true medicine'; Bodhi 2007 [1978]: 328–29). The

compound *tathāgata* is a reconstruction, a translation and Sanskritization of an underlying form, which is only imperfectly understood. It is surprising how little we know about one of the most common words of Buddhism; some scholars have even suggested that the compound is borrowed from a non IA source (Schayer 1935: 211–213; Thomas 1937: 186–187).

The Sanskrit word for ‘effort, exertion’, *pradhāna* (from S *pra* + *vdhā*, ‘devote oneself’), was pronounced in the earlier, underlying *koine* as *pahāṇ(n)a*, with the stop omitted and the aspiration alone remaining. *pahāna*, however, also meant ‘abandonment’ (from S *pra* + *vhā*, ‘forsake, abandon’) and the transmission shows several instances where the two meanings are confused and mistranslated, thus once again confirming the existence of the simpler *koine* form (Childers 1875: 157 sv *iddhipādo*; Levman 2011: 1–20).

Of course the underlying form was often not ambiguous at all, just a simplified, reduced form of the word. For example, the common S word *dirgha* (‘long’) appears in P as *digha* with the consonant cluster eliminated and in AMg as *dīha*, with the aspirated stop removed. Again, if one pronounced it *drigha* (as in G and the Dardic languages), *digha* (as in P), *dīha* (as in Amg), or *dirgha*, the common denominator to them all would be *dīha*, which has been preserved as the AMg form.

Sometimes only one transmission tradition has survived (usually P), but variant readings in the different Burmese, Sinhalese and other recensions reveal the potential underlying *koine*: *maggajjhāyi* (‘concentrated on the path’, Sinhalese) and *maggakkhāyi* (‘teacher of the path’, Burmese) in variant passages of the Sn indicated a translation from **maggahāyi* (Sn v. 85-b). In another example from the early Buddhist poetry, there are three cognate variants from the P, S and G transmissions: the different aspirated stops in P *palighaṃ* (*Dhammapada* 398-c, ‘cross-bar’) and S *parikhaṃ* (*Udānavaraga* v. 33.59A-c, ‘moat’) are probably due to a *koine* transmission similar to G *phali’a* (*Gāndhārī Dhammapada* [GDhp] 42-c) where both the aspiration and the stop have been lost (with a compensatory aspiration of the first stop *ph-*). The differential transmission of cognate forms *pramuñcantu* (‘let them free’, P) and *pramodantu* (‘let them rejoice’, BHS) and *praṇudantu* (‘let them reject’, BHS) appearing in parallel passages of the *Brahmāyācanasutta* are derived from a common source where the intervocalic stop has become a glide or disappeared, and the nasal is ambiguous (Levman 2012: 35–54). Often the *koine* is clearly identifiable, but its meaning is not, like P *nekkhama* (‘having left home’) – is it a simplification of *naiṣkramya*, ‘departure from the world’ (< S *niṣ* + *vkram*, ‘to go out, depart, leave worldly life’), *naiṣkāmya*, ‘without desire’ (< S *niṣ* + *kāma* ‘desireless, disinterested, unselfish’), or *naiṣkarmya* (< S *niṣ* + *karman*, ‘without action, tranquil’)? Or was it a pun meaning all three or two out of three? – certainly the last meaning seems most applicable to the Jaina tradition, where the word is also

found (as AMg *ṅikkamma*) and understood as ‘quiescent, free from karma’ the elimination of which is the ultimate Jaina goal (Mylius 2003: 318).

5 The language picture at the time of the Buddha

We have very little information about the linguistic fabric at the time of the Buddha and earlier. We may extrapolate – from the evidence of the Aśoka dialects of the mid third century BCE, – that in addition to the liturgical language of OI (Vedic), there were at least three principal Middle Indic dialects present in the fifth century BCE (east, west and northwest) and there may well have been others, now lost. There were also other Indo European languages present (like the ancestors of Tocharian and Krorainic),⁵ with different phonological structures to MI (e.g. no phonemic voiced stops) which further complicated MI’s linguistic development; although Tocharian and Krorainic do not enter linguistic history until the early centuries CE, their roots are nevertheless traceable to at least the second millennium BCE or earlier, and proto-forms of these may well have been influential on MI phonology (see below). In addition, the linguistic tapestry is further complicated by the presence of three or more pre-existent languages, native to the Indian sub-continent – Dravidian, Munda and Tibeto-Burman – and perhaps others not yet identified (Southworth 2005: 65).⁶ Emeneau characterized the Indian sub-continent as a “Sprachbund” or linguistic area, a term for an area in which “languages belonging to more than one family show traits in common which do not belong to the other members of (at least) one of the families” (1978: 201). In Emeneau’s definition of Sprachbund with respect to South Asia, the common traits belong to the Indo-Aryan languages (OI, MI, New Indo-Aryan), Munda and Dravidian but are not shared by Indo-Aryan’s closest cousin, Iranian.⁷ The mechanism which creates these shared features is extensive bilingualism, resulting from the interaction and

5 Krorainic is the language of the Niya documents from the Kingdom of Shan-Shan or Kroraina in central Asia; see Burrow (1937: v–ix).

6 I use the word “indigenous”, “native” and “autochthonous” to refer to those languages which were present in the Indian sub-continent before the arrival of OI/MI speakers. The word “pre-existent” is a better term which I use when possible; however at times the other words fit the context better. The use of these terms is not intended to allege that the pre-existing languages *originated* in the sub-continent, as they themselves were probably immigrant languages at some earlier time. Needless to say, this is a very complex issue, beyond the purview of this article.

7 See Southworth (2005: 88–90). Other languages in this Sprachbund include Tibeto-Burman and one or more other heretofore unidentified languages, like Witzel’s “Para-Munda” and Proto-Burushaski and the language of the Indus Valley civilization.

intermarriage of the Indo-Aryan speaking immigrants and the local population. Emeneau calls the process “Indianization” of the Indo-European (IE) component in the Indic linguistic scene (1956: 7); that is, borrowing of certain lexemic, phonological and structural elements from the pre-existent languages (Levman 2013: 147–152, Levman 2014: Ch. 11, 495–516). Although Emeneau’s Sprachbund hypothesis is not universally accepted, it is well supported by further research since his initial hypothesis, and the results of this study (see Kuiper 1967; Southworth 1974; Witzel 1999; Southworth 2005, for supporting views; opposing views are discussed in Bryant 2001: 76–107). Nevertheless whether the pre-existent languages affecting the development of IA are viewed as substrata (which were subsumed and displaced by the incoming Indo-Aryans), adstrata (geographically adjacent language groups) or superstrata (where Dravidian, Munda, etc. were themselves the “intruders”), makes no difference to the argument in this paper, which does not seek to prove or disprove an “out of India” hypothesis, but to demonstrate inter-linguistic structural and phonological interference.⁸

The linguistic fabric at this time was extremely complicated and its complexity is inversely proportionate to the amount of data we possess. The earliest Buddhist transmission was only preserved in a few dialects: P which is a mixed dialect with north-western, eastern and western elements; G, a north-western dialect; and BHS, the Sanskritized Prakrit of the Madhyadeśa (‘middle country’) or central north part of India. No Buddhist teachings have been preserved in Māgadhī or AMg, although the Jaina canon, written in the latter dialect, has a few renditions of gnomic wisdom passages which are also present in Pāli. By comparing cognate parallel passages in the surviving witnesses, we can uncover the existence of an earlier language of Buddhism which may be characterized as a homogenized, “lowest common-denominator” dialect with all major dialect differences removed or simplified and only the common elements amongst the dialects retained (Levman 2014); that is, by definition, it is a *koine*, a trans-regional common language that arose to facilitate interaction, communication and trade between peoples of diverse ethnic and linguistic background. One might characterize it as a compromise variety which retained those features easily recognizable to most speakers while eliminating those which impeded mutual intelligibility (Mufwene 1985: sv *koine*). It may well have arisen centuries

⁸ I in fact do not subscribe to the “out of India” hypothesis and hold with Witzel (2001) who considers it contradictory and unscientific. I will therefore continue to describe the Munda, Dravidian and Tibeto-Burman languages as “pre-existent”, “autochthonous” or “indigenous” throughout this paper (footnote 6). See Southworth (2005: 65), Figure 3.1 for an approximate distribution of these substratum languages before the IA migrations.

before the Buddha was born and been used by the Buddha himself and/or his immediate disciples for teaching. The structure and phonological nature of this common dialect was affected by several different factors:

- (1) The normal diachronic change of language over time tending to diversification, that is, the change of OI to MI which began in the late second millennium BCE and continued to develop throughout the first millennium resulting in the emergence of several new dialects. The first historical snapshot of this divergence occurred in the third century BCE with the Aśokan edicts; however, a study of Vedic phonetics shows “the extensive influence of Middle-Indic phonetics in the earliest periods of the language” (Bloomfield and Edgerton 1932: 20). Certainly by the time of the earliest G documents (first century BCE, the earliest Buddhist manuscripts we possess) or earlier, the dialects had diverged to a point where they were no longer mutually intelligible without requiring some form of translation or dialect levelling (#2).
- (2) A harmonization of the different dialects in a MI *koine* or common language which was intelligible across all the dialects. When this inter-language developed is impossible to say, but the linguistic evidence in this paper suggests that the earliest teachings of the Buddha were either composed in such a language or translated into it from the time of the Buddha’s teaching (fifth century BCE) or shortly thereafter. Although we have no independent historical “proof” of the existence of such an inter-language, it has been posited by several scholars as a logical necessity for both government administration and trade purposes (Geiger 1916: 3–4; Smith 1952: 178; Lüders 1954: 8) and the existence of other nearby, coeval inter-languages (Imperial Aramaic and Greek *Koine*) support the hypothesis.
- (3) The synchronic influence of the pre-existent languages (Dravidian, Munda and Tibeto-Burman) on the development of MI and its *koine*; because of their very different phonological limitations, speakers of these languages constrained and accelerated the evolution of MI in certain directions consistent with the phonology and phonetics of their mother tongues.
- (4) The potential effect of the two other influential interlanguages in common use from about 600 BCE onward – the Imperial Aramaic *lingua franca* and Greek *Koine* – as emulating models (for discussion of which, see below).
- (5) The potential influences of other languages, Indo-Aryan or Indo-European, which shared a different phonological structure to MI; at the time of the Buddha northern India was a major cosmopolitan crossroads and linguistic convergence centre mid-way between the Mediterranean world and China, and an important trading centre for the sub-continent in its own right.

- (6) Sociolinguistic factors which may have influenced the structure of the *koine* through emulation of the preferred north-western dialect, presumably the dialect spoken by the immigrant Indo-Aryans. Pāṇini was himself a north-westerner and it is this dialect which he established as the standard in his grammar; the dialect of the eastern tribes was considered inferior to the purer speech of the north and north-west.⁹ The *koine* is in fact most similar to G (but not identical with it; Levman 2014: 53, 461). I do not pursue sociolinguistic issues further in this paper as this is a separate discipline in and of itself.¹⁰

6 Characteristics of the *koine* and its development

The nature of this common dialect or *koine* I have already discussed in detail in previous work (Levman 2014; for summary, see Chapter 14: 637–642). The purpose of this paper is to examine how and why the *koine* developed in the form it did. I will argue that this is in part due to the attempt to harmonize diachronic forces tending to diversity, and in part due to the phonological constraints of indigenous non-IA speakers, who were at least initially – when the Indo-Aryans first entered India in the late second millennium – the majority of the population and who, in learning the increasingly dominant MI as a second language, had to adapt the pattern of their native phonology to the very different MI system. This catalyzed and accelerated the development and shape of the MI common dialect in ways which we will discuss below. We will start by examining the major characteristics of the *koine* and then look at the potential influences of the other languages on its form.

⁹ See for example, statements in the *Kauṣītaki-Brāhmaṇa* that those who want to learn the best speech go to the north (west), since the best known speech is spoken here, in Keith (1971 [1920]: 387). Also Oldenberg (1882: 400), notes: “With the Buddhists the capital of the Gandhāras, Takkaśilā, figures constantly as the place to which anyone travels, when he desires to learn something good”. See also Deshpande (1979: 254) where the non-Aryans (from the east) are accused of being *mṛdhra-vācaḥ* (‘with obstructed speech’).

¹⁰ For some recent work on sociolinguistic tensions between the use of S and Prakrit, see Deshpande (1993: 1–16); Levman (2008–2009: 33–51) who argues that the Buddha prescribed his teachings to be transmitted, not in S, but in *sakāya niruttiyā* (‘in my own terms/expressions’); and Kulikov (2013: 59–91) on the influence of MI vernacular on the formulation of Pāṇinian grammatical rules.

6.1 Lenition of intervocalic stops

In MI, the distinction between voiceless and voiced consonantal stops is phonemic (changes meaning), and can also mark dialect idiosyncrasies. Lüders believed that the language of early Buddhism – contained in the so-called “Urkanon” – weakened voiceless stops and eliminated voiced stops intervocalically. This accounted for the confusions that resulted when the underlying language was translated into Pāli; hyperpālisms resulted when the translator incorrectly restored a voiced consonant to its “original” voiceless state, or substituted the wrong intervocalic consonant for what was transmitted as a glide (Lüders 1954: §122–148). Indeed we have lots of examples of these confusions in P (and the other Prakrits) and there was indeed a dialect – or *koine* – where at least some of the intervocalic consonants had developed into a *-y-* or glide (Norman 2006: 85), whether through diachronic or synchronic forces (or probably both), we will discuss below. Another philologist, Mehendale (1968: 56), argued that these errors came about through borrowing, that the eastern source language of the teachings was characterized not by voicing but by devoicing and Pāli simply imported these without changing them back to their “original” state. However a clear pattern in P voicing and devoicing cannot be discerned, which is what one would expect if the pattern was a feature of P proper. A study of dozens of MI words reveals no clear pattern of voicing and devoicing attributable to dialect idiosyncrasies (Levman 2014: 475–493); it is a random phenomenon where identical phonetic environments show opposite voicing, suggesting that linguistic diffusion, that is, the effects on MI of other languages which lacked the voiced/voiceless distinction may be responsible for the confusion.

There is an additional weakening of intervocalic stops in G which shows a softening of the intervocalic voiceless dental stop (*-t-*) and the voiced dental aspirated stop (*-dh-*) to a fricative (written *-s-*, sounded [z]), which is similar to Aramaic and Dravidian (see below); this also occurs with the consonants *-k-* and *-g-* which are sometimes represented by *-ḡ-*, phonetically representing [ɣ], a voiced velar fricative (in the GDhp and the Niya Documents, Brough 1962: §31; von Hinüber, 2001: §173). This phenomenon crops up occasionally in the other Prakrits (e.g. S *antatas* > BHS *antaśas*, ‘so much as’) and manifests in Pāli with a change of the aspirated stop to the labiodental fricative *-v-* (e.g. *dhīro* > *vīro*, Norman 2006: 157). Historically the lenition of intervocalics might be expected to follow the pattern voiceless > voiced > spirantization > glide > Ø, but the development is not so clear-cut, as we find examples of all the changes in G and first examples of disappearance of intervocalic consonants or semi-consonants occur as early as the Vedas (late second millennium

BCE).¹¹ As Kenneth Roy Norman has noted (1993: 91), “we must recognise that there is not necessarily any correlation between the stages of linguistic development of kindred languages, so that two related and contemporary dialects can show vastly different stages of development, with one showing a far later stage of development than the other.”

6.2 Aspirated stops

We also know, because of later ambiguities in their translation, that aspirated stops were transmitted as aspirates only in the earlier language. So a translator, confronted with a word like *paha* (*Kevaddhasutta*, DN 1, 223¹²) did not know whether it referred to *S prabhā* or *patha*, or *prthak*, etc. He or she had to guess and the guess was not always right, as we can see from the confusion in the commentary on that particular word (Levman 2014: 378–387). OI and MI are unique in having ten aspirated stops (*kha*, *gha*, *cha*, *jha*, *ṭha*, *ḍha*, *tha*, *dha*, *pha*, *bha*), a feature which was not shared by most of the other languages that made up the south Asian linguistic scene. Lubotsky (1995: 140–141) makes a case that the change of *dh* > *h* in the Vedic texts was rule based, while admitting that *bh* > *h* was the result of dialect borrowing. The *communis opinio*, however, seems to follow Pischel (1981[1900]: §188) that the change of aspirated stops to aspirate only in the Prakrits (except for the palatal and retroflex ones) is a dialect phenomenon, although they are usually (but not always) restored in the P translation (von Hinüber 2001: §184).

6.3 Assimilation of consonant clusters

OI has a rich assortment of consonant clusters, the vast majority of which were assimilated in Middle Indic or resolved with the addition of an epenthetic vowel (especially in eastern dialects). Assimilation and/or resolution is also a common feature of other non IA languages.

¹¹ One of the earliest examples of stop lenition/disappearance being the S word *maireya* (‘intoxicating drink’; reflexes in P *meraya*, AMg *meraga*, Prakrit, *maïrea*), with cognate Vedic *madirā* (same meaning) pointing to a derivation from **madireya*, with *-d-* > \emptyset (von Hinüber 2001: §170). Another example is *pra-uga* (‘forepart of the shafts of a chariot’), derived from *pra-yuga* (Wackernagel 2005 [1896]: vol. 1, §37b). The earliest datable examples we have are from the Aśokan edicts (Levman 2010: 65), e.g. S *kādamba* > *kāamba* in Pillar Edict 5; *mama* > *maa* in Rock Edict 5 (Shābāzgaṛhi and Mansehrā); *devānāmpriaysya* > *devanapiasa* in Rock Edict 1 (Shābāzgaṛhi); S *iha* > *ia* in RE 13 (Shābāzgaṛhi). The phenomenon is also common in G, e.g. S *pratyaya* > G *prace’a* in GDhp 88; S *bhoga* > G *bho’a* in GDhp 261; S *makṣikā* > G *makṣi’a*, GDhp 285, etc.

6.4 Levelling of sibilants

G is the only MI Buddhist transmission dialect that we know of which maintained a distinction between dental, palatal and retroflex sibilants; the other dialects – and the *koine* – levelled them to one dental (s) or palatal (ś) sibilant (in the case of Māgadhī).

6.5 Interchange of glides with glides, glides with nasals, glides with palatals and liquids

In MI *-y-* and *-v-* were often interchangeable, as were *-y-* and *-j-*, *-v-* and *-m-* in nasalized contexts; some of this interchange was due to MI dialect idiosyncrasy (or inherited from OI, cf. Bloomfield and Edgerton 1932: §223–240), while it may also be in part attributable to the lack of a *-v-* sound in some non IA languages like Munda, Tibetan (Tib) and Chinese. The phonemes *l* and *r* were also interchangeable, usually thought to be because of dialect differences with *l* predominating in the east of India and *r* in the west.

7 The Greek and Aramaic *Koines*

These are the principal simplifications in the underlying common language or *koine*. In what follows (Section 8) we will look at some of the phonological constraints in other Indic languages that may have contributed to this structure. But first we will look at the theory of how a *koine* develops and then examine the two other *lingue franche* which were prevalent at the time of the Buddha – Aramaic and Greek *Koine* – to increase our understanding of the structure of inter-languages, and also to examine possible reciprocal influences between these and the MI *koine*.

7.1 How a *koine* develops

A *koine* results from dialect levelling and simplification, primarily due to (1) the elimination of interdialect phonological differences which impede understanding, and (2) the structure and influence of the surrounding native languages, whose speakers had to learn a foreign language and communicate with the foreign speakers. Modern studies have shown that in face-to-face interaction between speakers of different dialects, speakers accommodate to

each other linguistically by reducing the dissimilarities between their speech patterns and adopting features from each others' speech (Trudgill 1986: 39). Accommodation is the reduction of pronunciation dissimilarities through a) alternating one's own variant of a form with that of the other speakers; (b) using the other speakers' variant in some words but not others (transfer/mixed dialects); and (c) using pronunciations or forms intermediate between those two in contact. Of course, all three may occur in conjunction with each other. (Trudgill 1986: 62). By reducing differences between speech forms a *koine* or common language results, which is a common denominator language where all dialect idiosyncrasies have been removed. So, for example, a word like *dirgha* in S (discussed above), with an aspirated stop and conjunct becomes *diha* in the common language as already discussed, with the loss of the consonant cluster *-rgh-* and the change of the aspirated stop to an aspirate only). Thus the distinctive feature of G, which preserves the *-r-* and metathesizes it (*drīgha*) is omitted, as is any dialect where the voiced vs. voiceless consonant distinction is important (as the stop has disappeared). So one often finds in the underlying common language that all stops are sometimes eliminated intervocally and replaced by a glide, which allows the hearer to insert the correct stop according to the phonology of his/her dialect and his/her understanding of the context. This accounts for how we get different interpretations of parallel cognate passages in the DhP, like *pāceti* ('bring to an end'), *prājeti* ('drive forward') and *prāpayati* (= *prāpeti*, with *-aya-* > *-e-*, 'to lead') with *payedi* underlying (the GDhp form being in this case the same as or similar to the *koine*); *sahavya* ('friendship') or *svabhāva* ('nature, condition'), with **sahāva*, **sahāya* or **sahāa* underlying; or *virajo* ('stainless') and *virato* ('ceased') with *virayo* underlying, to name a few examples.¹² This is what is called the loss of marked forms. There are also other morphological simplifications which take place in a common language, like an increase in morphophonemic regularity, increase in invariable word forms, symmetrical paradigms for declensions and conjugations, etc.

Contact with different dialects is one major influence of *koine* formation; the other is contact with different languages, especially languages which may have a different phonological structure than MI; this is only a difference in degree, not in kind, for both forces (differences in dialects and languages) act to shape the common language through interference. In his classic study, *Languages in Contact* (first published 1953), Uriel Weinreich lists four phenomena of interference between two phonemic systems which come into contact (1967 [1953]: 18–19):

¹² For discussion see Levman (2014: 79–81; 245–247; and 401–403).

- 1) Under-differentiation, that is, not distinguishing phonemes in the new (immigrant) language which are lacking in the primary language. For speakers of native Indic languages (like Dravidian and Munda) learning MI, they would not hear aspirated stops phonemically, for example, which are not present in their native language.
- 2) Over-differentiation of phonemes involves the opposite, that is, imposing phonemes from the primary system on the new language. A good example of this phenomenon is the introduction of the retroflex system into IA, which is believed to have been introduced by native Dravidian and/or Munda speakers learning IA as a second language; they recognized allophones in OI which were close to the Dravidian/Munda retroflex system and assigned them to these retroflex phonemes which were eventually assimilated into IA (Emeneau 1974: 93; Deshpande 1979: 297; for discussion see Levman 2014: 504–505; for opposing view Hock 1996). Particularly noticeable in this respect are loan words from Dravidian and Munda where no attempt was made to conform to IA phonology (e.g. most IA words in *-ṇḍ-*, like *cāṇḍala*, *muṇḍa*, etc. See Woolner 1926–1928; Mayrhofer 1963: vol. 1, 370, vol. 2, 651).¹³
- 3) Reinterpretation of distinctions occurs when the bilingual speaker interprets redundant, incidental features in the new language as significant because of their relevance in his/her own phonological system. This is a form of over-differentiation which could work in either direction. A native Dravidian speaker, hearing an alveolar stop (/r/) as an allophone of a dental stop, may interpret it phonemically; a native MI speaker, hearing an allophonic intervocalic stop pronounced by a Dravidian speaker might also consider it semantically meaningful, although it is not in Dravidian.
- 4) Phone substitution applies to phonemes which, though identically defined in both languages, are pronounced differently. This is especially relevant with vowel sounds; for example, both Dravidian and MI have short and long vowels, but a small, idiosyncratic difference in pronunciation could easily cause confusion.

In addition to the above, there is also the “complicating possibility” of hyper-correction that may take place both in listening and in speech (1967 [1953]: 19), where overcorrection takes place because of misunderstanding on the part of the interlocutor.

¹³ Southworth (2005: §3.31) suggests that Munda (which only preserves the retroflex *ḍ* (see Table 4) and Dravidian may themselves have obtained the distinction from an earlier substrate language, i.e. the language of the Indus Valley civilization. See also Witzel (1999: 14): “In short, the people of the (northern) Indus civilization must have spoken with retroflexes”.

We will see that most of the factors influencing the form of the common language relate to item #1 above, under-differentiation, which leads to simplification by eliminating the ambiguous sound from the speech-form. Also important to understand is the fact that it is the immigrant, rather than the indigenous language which is most subject to interference. Weinberg gives the following three reasons for this phenomenon (1967 [1953]: 91):

- 1) The immigrant language must adopt indigenous vocabulary for new flora, fauna and other unfamiliar phenomena they encounter. In fact, we can identify dozens of toponyms and names of biota and cultural practices which were adopted by the IA migrants (Levman 2013: 148–149).
- 2) Borrowing from the new language provides the socially and culturally disoriented newcomers with some familiarity and stability, thus weakening their natural resistance to compromising the “purity” of their own language.
- 3) The necessity of intermarriage, because of the lower proportion of women amongst the immigrant population. Recent genetic studies, for example, suggest that ancestral South Indians spoke a Dravidian language and Y chromosome and mtDNA (mitochondrial) analysis shows a significant male gene flow from groups with more ancestral North Indian relatedness into ones with less; that is, male Indo-Aryans taking Dravidian wives (Reich et al. 2009: 493).¹⁴

7.2 Aramaic

Aramaic is a Semitic language, closely related to Hebrew, both members of the Northwest Semitic group. The language was widely spoken in the late second millennium and throughout the first millennium BCE in the Near East (from present day Turkey to Iraq), and was adopted around 600 BCE as the official language for the eastern Persian empire, for communication between peoples of different language backgrounds – i.e. as a *lingua franca*. As is well known, the Persian emperor Darius took control of the Indian sub-continent north of the Indus River in the late sixth century BCE, and governed it until he was defeated by Alexander. Generally termed “Imperial” or “Official” Aramaic, it was still used in this form until about 200 BCE, despite the demise of the Persian empire in the fourth century. Aramaic was written in a Phoenician based script (right to left) similar to the Kharoṣṭhī of the GDhp and other texts (Salomon 1998: 25). Jesus was a native speaker of Aramaic and various dialects of Aramaic remained the regional *lingua franca* of the Near East until displaced by Arabic in the seventh century CE.

¹⁴ Thanks to Prof. Alexei Kochetov for this reference.

7.2.1 Phonology

See Table 1 for a list of Aramaic consonants. What distinguished the Imperial Aramaic *lingua franca* from Early or Old Aramaic (950–600 BCE) are the following simplifications (from Segert 1997: 118–125):

- (1) All the interdental spirants merge with the dentals; the spirants cease to be phonemic and become allophones of the dentals (cf. G dental > sibilant change noted above).
- (2) The uvular consonants disappear and merge with the pharyngeals (*ġ* or IPA [ɟ] > ‘or IPA [ʕ]; *ħ* or IPA [χ] > *h* or IPA [h]). In modern Hebrew the pharyngeals merged with the velars as most Ashkenazi Jews could not pronounce them (Kerswill and Williams 2000: 71).
- (3) the glottal stop (ʔ) was elided at the end of words and syllables.
- (4) regressive total assimilation of the nasal /n/ to the immediately following consonant is very frequent (e.g. *-nd-* > *-dd-* or *-nt-* > *-tt-*).
- (5) semivowel *y* can be elided between two long vowels (*qāyēm* > *qāēm*, ‘standing up’).
- (6) short vowels in open syllables were reduced or elided (e.g. *malkatā* > *malk²tā*, ‘the queen’). *t* = IPA [θ].
- (7) new vowels are inserted to avoid clusters of consonants: *malk* > *mælxæk* > * ‘king’ (epenthesis or anaptyxis). *k* = IPA [x].
- (8) compensatory vowel lengthening for loss of weak consonants or simplification of a doubled consonant.
- (9) monophthongization of /aw/ and /ai/ into /ō/ and /ē/ respectively.

Note that we see a lot of similarities between Imperial Aramaic and the MI dialects described above (assimilation or resolution of consonant clusters, dropping of final consonants at the end of a word, monophthongization of diphthongs). Although MI has no interdental spirants (θ or ð), in G at least, they were an allophone of the dentals in certain contexts (e.g. P *madhura* = G *masuru*, ‘sweet’ GDhp 54; P *vanathajā* = G *vaṇaše’a*, ‘born of craving’, GDhp 89); the intervocalic consonant *-th-* or *-dh-* was a fricative, which sounded close enough to [z] to be written as *-s-* or *-š-*.¹⁵ Was this also a feature in the MI *koine*? We certainly do find confusion between dental stops and sibilants in the received texts, suggesting pronunciation variation in the underlying oral tradition. Sometimes this can lead to semantic ambiguities in the transmission as in the case of parallel cognate

¹⁵ Brough (1962: §43 and §43a). He also hypothesizes that “the situation in the *Dharmapada* strongly suggests that the development of the earlier unaspirated dental stops to fricatives followed that of the aspirates, so to speak, one stage in arrears.”

Table 1: Aramaic Consonant Inventory (from Segert 1997: 119).

	Early Aramaic		Imperial Aramaic		Biblical Aramaic allophones	
	Emphatic	Voiced	Unvoiced	Emphatic	Voiced	Unvoiced
Laryngeals		<ʔ>	<h>		<ʔ>	<h>
Pharyngeals		<ʕ>	<h>		<ʕ>	<h>
Uvulars		/g/	/b/			[g]
Velars	<q>	<g>	<k>	<q>	<g>	<k>
Palatovelars			š		z	š
Sibilants	<š>	<z>	<s>	<š>	<z>	<s>
Interdentals	/z (t)/	/d/	/t/			[š] [t]
Dentals	<ṭ>	<d>	<t>	<ṭ>	<d>	<t>
Labials			<p>			<p>
Bilabials						[b]
Linguals	<l>	r		<l>	r	
Nasals	<n>	m		<n>	m	
Semivowels	<y>	w		<y>	w	

Note: Symbols: <> graphemes; //, reconstructed phonemes; [] phonetic pronunciation; underdot (e.g. ṭ), emphatic, i.e. pharyngealized [ṭ]; underline or overline (e.g. ṭ, ḡ spirantized [t] or [g] same as [θ] or [χ]). The author would like to thank the publisher Eisenbrauns, for permission to reprint this table which originally appeared in a 1997 publication *Phonologies of Asia and Africa* (page 119).

versions of the P *Sutta nipāta* v. 64-a, where the P has *ohārayitvā* ('having discarded' < S *ava/apa* + *vhr*, causative, 'to take down, put down, cause to throw away'), the BHS *Mahāvastu* (1.358⁵) has *otārayitvā* ('having thrown off' < S *ava* + *vṭr*, caus. 'take down, take off, turn away from'); and the G version (Salomon 2000: 146, verse 19-a) has *ośaḍaita* ('having cast off' < S from *ava* + *vśaṭ*, caus. 'cast off' or < S *apa* + *vśad* in caus., 'to cause to fall off or away'). Although the meanings are all similar, the verbs are all different. The underlying *koine* transmission was therefore either **o'ārayitva* (which is attested in AMg, *o'āra* = *avatāra*),¹⁶ where the intervocalic has been dropped and each tradent interpreted it differently, or **oSārayitva* (S = sibilant): taken as given in the G version (with a substitute of *-ḍ-* for *-r-* and loss of the intervocalic glide);¹⁷ or in the BHS Mvu, the sibilant taken as an allophone for a dental stop; or in P interpreted as a deliberate "Māgadhism", where a sibilant often morphs into an aspirate (for example gen. sing. *-asya* > *.*āsa* > *-āha*; Pischel §264; von Hinüber 2001: §221).¹⁸ However one explains the differences, there does seem to be some confusion over the sibilant/stop/aspirate relationship. A similar ambiguity is found in parallel versions of the P *Sattajaṭṭilasutta* (*Samyutta Nikāya* 1, 79¹⁰, and Udāna 66⁷) where the variant forms *osāpayissāmi/ oyāyissāmi/ ohayissāmi/ osārissāmi/otarissāmi* occur.¹⁹ Since we may assume that Aramaic was in use in northern India at the time of the Buddha as a common trading and/or administrative language (judging from the number of Aramaic rock inscriptions that have survived; Norman 2012: 43–44), did that language's merger of the interdental spirants with the dental stops affect their pronunciation in MI?

7.3 Greek *Koine*

The word *koine* originally denoted the common literary dialect of the Greeks (*ἡ κοινὴ διάλεκτος*) from the close of classical Attic to the Byzantine era (OED). It was the established language of commerce, diplomacy and officialdom from at least the reign of Alexander's father, Philip (360/59–336 BCE) while the Atticization of the Macedonian court had begun a century earlier (Horrocks 2010: 80–81). Although the use of Greek *koine* in the Indian sub-continent probably post-dates Aramaic and MI *koine*, it is nevertheless instructive to

¹⁶ Pischel (1981 [1900]: §154), hereinafter "Pischel", and Mylius (2003: 186).

¹⁷ Pischel §241, 258.

¹⁸ There are other possible derivations as well, but I have only given the two most obvious.

¹⁹ See discussion in Levman (2014: 350–54). For a more detailed discussion of the first case, see also pages 250–55 in the same work.

examine its structure, which has some similarities to the other inter-languages. Whether this is due to influences from Aramaic and/or the MI *koine*, or linguistic universals is impossible to determine.²⁰ Like Aramaic and other *koines*, the “original” Greek *koine* is a “compromise variety”, consisting of “features easily recognizable to speakers of most Greek dialects and dispensing with those that most often impeded mutual intelligibility” (Mufwene 1985: s.v. *koine*). The principal simplifications that occur in Greek *Koine* are (after Colvin 2007: 67):

- (1) Vowel length is lost. For a Greek Koine speaker, this might lead to under-differentiation of MI vowels where length is phonemic (above §7.1.1). This is exacerbated by the fact that in early Kharoṣṭhī and Brāhmī script vowel length is not usually notated (Brough 1962: §20; Norman 2006: 107).
- (2) The monophthongization of several diphthongs. Similar to Aramaic and actualized in all the Prakrits.
- (3) The voiced stops β, δ γ, become fricatives [v], [ð] and [ɣ]. Similar to G as discussed above.
- (4) Aspirated stops φ [p^h], θ [t^h] and χ [k^h] are lost and become fricatives [f], [θ], [x]
- (5) The affricate/cluster ζ becomes a simple voiced fricative [z]. In most Prakrits the affricate *kṣ* > *ch* or *kh* (Pischel §317); In G it was pronounced as a retroflex fricative [ʂ] (Bailey 1946: 774) which is corroborated by Chinese transliterations (Levman 2014: 579).
- (6) The aspirate or rough breathing before a vowel disappears (psilosis or de-aspiration).
- (7) Final *-n* becomes weak or non-existent.

Note the common fricativization of stops in Greek *Koine* which also occurs in Aramaic and in MI dialects. The language of Shan-Shan (Krorainic, an IA language), for example, was devoid of voiced stops, and the voiceless stops were voiced and spirantized (weakened to fricatives) intervocalically (Burrow 1937: §16). More on this below. Krorainic also lacked aspirates, so that their appearance and disappearance was sporadic (§24). This randomness is also characteristic of other MI dialects, like P (Geiger 2005 [1916]: §40, hereinafter “Geiger”), probably constrained by the lack of such consonants in the native languages, and, as we have seen in G, and the other dialects, there are lots of examples where an aspirated stop changes to a fricative (*dhīro* > *vīro* above).

²⁰ We do have historical evidence that in parts of the Indic sub-continent both Greek and Aramaic were common enough to warrant translation of Aśoka’s edicts into these languages. There are two Greek inscriptions which have survived in Kandahār, Afghanistan: one bilingual Greek and Aramaic similar in content to minor Rock Edict 1, and one translating portions of Aśoka’s Rock Edict 12 and 13 (Norman 2012: 43).

8 Other languages

8.1 Dravidian

Dravidian is one of the autochthonous languages of the sub-continent, whose presence pre-dated the arrival of Indo-Aryans, by perhaps as much as a millennium.²¹ For the phonemic inventory of Proto-Dravidian (PD) consonants, see Tables 2 and 3. In the Proto-language there is no distinction between voiceless and voiced stops; they are in complementary distribution and voiced stops are allophones of their voiceless counterparts. Per Caldwell's law, stops are voiced when intervocalic or when they follow a homorganic nasal;

Table 2: Proto Dravidian consonant inventory (after Steever 1998: 14).

Labial	Dental	Alveolar	Retroflex	Palatal	Velar
Stop p	t	<u>r</u>	ʈ	c	k
Nasal m	n		ɳ	ɲ	
Lateral	l		ɭ		
Tap	r				
Approximant			ʐ		
Glide v				y	(h)

Note: Proto-Dravidian (PD) has 10 vowels, in five pairs, with each pair containing one short and one long vowel, a, ā, i, ī, u, ū o, ō, e, ē.
 ʐ = IPA [ʐ], voiced retroflex fricative.

Table 3: Proto-Dravidian (absence of sibilants, after Zvelebil 1990: 7).

*p	*t	*t	*ʈ	*c	*k
*m		*n	*ɳ	*ɲ(?)	
			*ɭ		
		*r	*ɻ		
*v				*y	
					*H(?)

²¹ Southworth (2005: 325) suggests that Proto Dravidian is probably contemporary with early Harappan culture (mid third century BCE).

elsewhere they are voiceless, including when they are geminate or initial (Steever 1998: 15). According to Zvelebil (1990: 7–8), Dravidian lax (voiced) obstruents are always weakened intervocalically, and in historical development the non-apical (alveolar or retroflex) stops are normally further weakened to voiced fricatives ($k > [ɣ, x, h, ç]$, $c > [dʒ, ʃ, ʝ]$, $t > [ð]$, $p > [β, v]$). When exactly this fricativization happened is not clear, although it was apparently a feature of Old Dravidian (Andronov 1968: 3). Note the similarity with fricativization in G, Aramaic and the Greek *koine*.

The alternation of *-m-* and *-v-* in Dravidian is another feature which it shares with G and other MI dialects (Zvelebil 1990: 10; Pischel §261). In Dravidian the change happens intervocalically as in G, as well as word initially.

PD also has no aspirated stops and is unusual in having no sibilants in its phonemic inventory. Wüst (1957: 83) ascribes this peculiarity to certain irregular correspondences in S and Prakrit like *śākinī* ('female demon') ~ *ḍākinī* ('female imp'), or *śākvara* ('mighty, powerful') ~ *ṭhakkura* ('deity, object of reverence') or *ṭhiṇṭha* ('gaming house') ~ Pāli *soṇḍa* ('addicted to'); presumably the sibilant was heard as a retroflex by a Dravidian bilingual speaker. The PD approximant /z/ has no exact equivalent in IA and when imported appears as /r/, /l/, /d/ or /l/. Adjacent stops were regularly assimilated in PD and if consonant clusters occurred it was usually only at morpheme boundaries (Steever 1998: 16). Some scholars have also argued that simplification of consonant clusters in MI was influenced or caused by Dravidian phonological patterns, in which consonant clusters are highly restricted and two stops very rarely occur together – the occurrence of two contiguous consonants is restricted to homorganic nasal plus obstruent, geminates, and liquid or glide plus obstruent, liquid plus geminate or liquid plus nasal plus obstruent (Andronov 1968: 3; Zvelebil 1990: 12).²²

PD is unusual in having a dental stop *t*, an alveolar stop *t*, (also written as *r*) and a retroflex stop *ṭ*; this would lead to over-differentiation on the part of a Dravidian bi-lingual speaker learning MI.

²² For discussion on the Dravidian influence on MI geminates, see Ananthanarayana (1991: 256). "Although reduction of consonant clusters to geminates can very well develop in a language without external pressure, the fact that Dravidian had only medial plosive geminates or sequences of nasal and stop may have contributed to the development of geminates in MIA. Similarly, the presence of single initial stops in Dravidian may have been responsible for the reduction of initial consonant clusters to single stops in Prakrits. It may be that the Dravidian bilinguals in Prakrits effected such changes since they were not used to consonant clusters in their own languages."

8.2 Munda

For the phonemic inventory of Proto-Munda (PM) consonants see Table 4. The Munda language family occupied the Indian sub-continent well before the arrival of the Indo-Aryans and the oldest stratum of loan words in the Rigveda is derived from Munda, or a related language designated as “Para-Munda”, whose genetic relationship to Munda is unclear (Witzel 1999: §1.2–§1.4). PM has an asymmetry in the stops with a voiceless dental *t* and a voiced retroflex *ɖ* and some therefore argue that it did not originally have a retroflex series, but that these were imported from an earlier, unknown substratum. It does not have a phonemic contrast between vowel length and lacks the glide *v* and the aspirated stops. At morpheme boundaries the contrast between voiced and voiceless stops is neutralized and they are replaced by checked sounds (phonetically unreleased and glottalized); so, for example, in present day Santali, *dak*’, ‘rain’ > *dag-a* or *dak’-a* before /a/, (‘it rains/will rain’). In this kind of situation *-k-* and *-g-* are allophones (Ghosh 2008: 31), which would lead to under-differentiation of intervocalic voicedness on the part of a Munda speaker learning MI. This lack of voicedness contrast is also reflected in the phenomenon called “rhyme-words” by Kuiper where there is free variation between voiced and voiceless stops word-initially and medially (Kuiper 1965: 59–66).

Table 4: Proto Munda consonant inventory (after Zide 1969: 414).²³

p	t	(c)	k				
b	ɖ	j	g	b’	ɖ’	j’	g’
	s						
m	(n)	ɲ		m n		ɲ	ŋ
	l			l			
	r			r			

Like Dravidian, in Munda languages, two stops almost never occur together (except at morpheme boundaries) and consonant clusters are always N + stop or stop + L (liquid) (Ghosh 2008: 31; Osada 2008: 103). This feature is confirmed for the earliest

²³ See also Pinnow (1959: 426–427) who reconstructs an “Uraustroasiatisch und Urmunda – Archiphoneme” which has no retroflexes but a voiceless and voiced (*t* & *d*) dental stop and a younger stage which he terms “Urmunda” which has both the dental stop contrast and the retroflex (*t* & *ɖ*). The former has no sibilant and the latter has a voiceless postalveolar fricative, *f* and also adds a uvular stop phoneme (*q* and *G*, where *G*=*h*). In what Pinnow calls the “youngest protolanguage stage (“jüngstes voreinzelsprachliches Stadium”), all variants are added which include the aspirates, checked consonants, and the interchange of velar and uvular stops and dental and retroflex stops.

stratum of the language (UrAustroasiatisch) which had a syllable structure OVO [(C)V(C)], where O = any consonant which could also be null and V = any vowel (Pinnow 1959: 457–458).

8.3 Tibetan

No one has yet succeeded in producing a definitive reconstruction of the proto-Tibetan language, a very daunting task considering the dozens of dialects that exist and the paucity of data available on many of them. A useful classificatory scheme – which may well be diachronically accurate (Sprigg 1972: 556) – identifies dialects as either cluster or non-cluster at word beginning; the former appear to be more archaic with a simple vowel system and with no distinction of length, nasality or tone. Non-cluster dialects (like Lhasa) have no word-initial consonant clusters but more complex vowel systems and tonal distinctions (Denwood 1999: 26). The simplification of word-initial clusters is consistent with the Lhasa dialect's position as a *lingua franca* interlanguage or *koine* in Tibet, where many of the dialects are mutually unintelligible. Although we are fairly confident that the Tibeto-Burman language group was one of the prehistoric languages of India, very little is known about their early history, except for their general location in the foothills of the Himalayas, where the Buddha was born (Southworth 2005: 65). Witzel suggests that the names of various Nepalese places (like Kosala, the kingdom which the Sakyans were vassals of), and various rivers and Himalayan tribes mentioned in the Vedas (Kāśī, Kirāta) came from a Tibeto-Burman substrate (1999: §2.5) and others have suggested that the Tibeto-Himalayan language family was influenced by an Austro-Asiatic Munda substrate (Sharma 2003).

Tib is a member of the Tibeto-Burman language family and an inventory of simple consonants is given in Table 5. Notice that in the proto language the contrast between voiceless vs. voiced stops is phonemic and there is no aspiration. We have no records of Tib written or spoken until writing was introduced in 650 CE, which recorded the pronunciation of what has come to be known as “Old Tibetan” (7th–11th centuries). Old Tib also has several peculiarities vis à vis IA. There is no distinction of vowel length (Hill 2010: 116). Aspiration is non-phonemic and all final stops are voiceless, even though written as voiced (2010: 119, 122). Old Tib ceased to exist with the collapse of the Tib empires and was replaced by Classical or Written Tib, the language of most Buddhist texts. Here aspiration is apparently phonemic (*ka* ‘pillar’, *kha* ‘mouth’) in word-initial position. Although there is apparently a phonemic contrast between voiceless and voiced stops in Classical Tib (Delancey 2003a: 256), for an IE trained ear, it is very hard to hear: the

Table 5: Simple consonants in Proto Tibeto-Burman (from Matisoff 2003: 15).

	Labial	Alveolar	Palatalized alveolar	Palatal	Velar	Glottal
Voiceless stop	p	t			k	
Voiced stop	b	d			g	
Nasal	m	n			ŋ	
Fricative		s, z	ś, ź			h
Affricate		t͡s, d͡z	t͡ś, d͡ź			
Lateral		l				
Tap or trill		r, ɾ				
Approximant	w			j		

Note: reconstructed Old Tibetan from Hill (2010: 121): k, g, ŋ, t, d, n, s, z, p, b, m, ts, dz, y, ʃ, r, l, l, ɰ, h, w, ɨ (palatalised front vowel).

difference between voiceless *-k-* and voiced *-g-* sounds much more like a high tone, low tone contrast than a voicing contrast, which by some is regarded as an allophone, at least in the dialect of Lhasa, where the difference is not phonemic (DeLancey 2003b: 270).²⁴ Hahn (2002: 12) lists five sets of stops (*p, ph, t, th, ʈ, th, c, ch, k, kh*) and the glottal stop (ʔ), none of which are voiced; voicing only occurs when there is a nasally prefigured medial consonant (e.g. ^mb, ⁿd, etc.), that is, in a specific phonetic context, therefore not phonemic. In the Tibeto-Burman languages of the Himalayan Region (including Ladakhi, a principal dialect of northeastern India), the phonemic contrast between voiced and voiceless stops seems to be neutralized in some dialects, for example, Tib *bumo*, ‘daughter’ = Kāgate, *po mo*, Sharpa, *pu mu* and Bhoṭiā (Sikkim) *pum*; classical Tib *-d* = dialect *-t* in some dialects: *bdun*, ‘seven’ > Kāgate, *tün/tin*; or *dos*, ‘load’ > Bhoṭiā > *toi*; classical Tib *-g* > dialect *-k* in Tib *brgyad*, ‘eight’ > Kāgate *ke*.²⁵ Tib is a monosyllabic language with a complicated phonotactic structure, but there are no consonant clusters within a word if the palatalized velars /kʲ/ and /khʲ/ and the prenasalized stops (in those dialects where they occur) are analysed as unitary segments (DeLancey 2003b: 272); between words, since the only permitted finals that are allowed are *-b*, (sounds as *-p*) *-l*, *-r* and the nasals (final *-g* > ʔ and final *-s* and *-d* modify the preceding vowel), consonant clusters with two dissimilar stops like IA never occur. Like the absence of most conjuncts in Dravidian and Munda, the Tib

²⁴ See DeLancey (2003b: 270) where the difference is not phonemic as Lhasa lacks a voiced stop.

²⁵ Bonnerjea (1936: §1, §3), leaving out the numerous cases like Tib *mig* > Ladakhi *mik*, which are presumably orthographic (as these both end in glottal stop).

phonotactic structure may also have been a factor in precipitating the assimilation of IA conjuncts to geminates, so that bilingual speakers could both understand and speak the language more easily. Of course we have no idea if this kind of Tib *koine* characteristic of Lhasa was in use in the Buddha's day, over a millennium earlier. That there was an ancestor language to Tib at the time of the Buddha, we can be certain, and it is reasonable to hypothesize that that language would have had some of these phonological features, though which ones ancestral and which derived is impossible to tell. So though we may not be able to pinpoint the diffusory effects of ancestral Tib on MI, it is logical to assume that it had some influence. Since medieval times the Lhasa dialect has acted as a *koine*, an inter-language levelling the many inter-dialect differences, and similar to some of the phonological simplifications in Aramaic, Greek and MIA; its inclusion here is relevant for that reason alone.

8.4 Tocharian

The Tocharians were inhabitants of the medieval city-states on the north perimeter of the Tarim Basin, right on the silk route of north-west China. Tocharian A is generally associated with the city of Agni (and therefore also called Agnean or East Tocharian) and Tocharian B with Kucha or West Tocharian). The documents that have survived are fairly late – from the sixth to eighth century CE and are almost all translations of Buddhist texts, but their language heritage certainly goes back several millennia to prehistoric times, although no one is quite sure who their ancestors are – perhaps to the Afanasievo culture to the north (3500–2500 BCE; Mallory and Mair 2000: 294–296), the Bactria-Margiana region to the west (2100–1900 BCE; Witzel 1999: 54) or the Qāwriḡhul culture of the second millennium BC in Chinese Turkestan (van Driem 2001: 1064). Indeed there are several words with an apparent Tocharian pedigree which are found in Vedic or Avestan, which may be explained on the basis of a Central Asian substrate assimilated into the Vedic writings.²⁶

Tocharian is an Indo-European *centum* language, retaining the velar stop /k/ of Proto IE where this sound was changed to an alveolar fricative [s] in Avestan, OI, MI and other surrounding languages; it is therefore an

²⁶ See Witzel (1999: 54–56): *kha-ra/xara*, ‘donkey’, cf. Toch. B. *ker-ca-po; iṣṭi, iṣṭikā/iṣṭiia*, ‘brick’, cf. Toch. *iṣcem*, ‘clay’? **medh/melit*, ‘sweet, honey’, IE **medhu*, Vedic *madhu*, Avestan *maḍu*, cf. Toch. B *mit*, ‘honey’, *mot*, ‘intoxicating drink’. This latter word may well be a joint inheritance from a common IE source. “In short, western and central Iran must have been inhabited by (archaeologically well attested) peoples of non Indo-Iranian speech.”

anomalous IE language surrounded by *satem* languages. The Tocharian consonant inventory is given in Table 6. For our purposes the most significant aspect phonemically is that there are no voiced stops or aspirated stops, meaning that speakers of Tocharian (or of Proto-Tocharian, for which this was also true) would not have heard these as meaningful phonemic differences – *k*, *g*, *kh* and *gh* would all sound as *k*, for example. There is also no distinction between long and short vowels. Although we do not know how widespread Tocharian was as a language during the time of the transmission of the Buddha’s teachings, its presence along the silk route suggests its potential importance and influence. Some of the confusion on voiced vs. voiceless stops in MIA may be attributable to Tocharian or Proto-Tocharian speakers who had to learn MIA without the benefit of the voicing/voiceless distinction in their aural inventory, thus making random mistakes in the audition and notation. As in Dravidian, intervocalic consonants would be allophonically voiced (as were stops before a nasal), thus leading to potential confusions with MIA dialects where the voicing was phonemic. For example, the word for S/P *nāga* (‘serpent, elephant’) in Tocharian is *nākās* with a *-k* instead of *-g* (which might be mistaken for S *nāka*, ‘firmament’), and the word for Buddha is *pa* or *pūd* (poetic) or *putti* (= S *pūti*, ‘purity?’).²⁷ Tocharian also had no *h* sound (voiceless glottal fricative) like Munda, or *v*, like Tibetan, Munda and Chinese (Table 8).

Table 6: Tocharian consonant inventory, (after Krause and Slocum 2007–2010).²⁸

	Labial	Dental	Alveolar	Palatal	Velar	Labiovelar
Stop	p	t			k	kʷ
Affricate		ts	c			
Sibilant		s	ʃ	ś		
Nasal	m	n		ñ	ŋ	
Liquid		l	r			
Glide				y		w

²⁷ The word is actually a compound *pañākte*, where *nākte* is an adjectival derivative of *~nakte* meaning ‘god’ (<http://ieed.ullet.net/tochB.html#pan%CC%83a%CC%88kte>). *putti* occurs in the compound *puttiśparām*, which is a noun meaning ‘Buddha-dignity’. See <http://www.utexas.edu/cola/centers/lrc/eieol/tokol-1-X.html#L211>. The letter *-ā-* denotes a mid high front vowel (accessed December 2014).

²⁸ http://www.utexas.edu/cola/centers/lrc/eieol/tokol-1-X.html#Tok01_GP01_02. (accessed December 2014).

8.5 Krorainic (language of the Niya documents)

Krorainic is an IA language which devoices voiced consonants and is usually grouped with G in the IE classification system (Cardona 1985: vol. 22, 618; Masica 1991: 52). It was the native language of the Kingdom of Shan-Shan on the south side of the Tarim basin, not far (as the crow flies) from Kucha and Agni, the inhabitants believed to be ethnically and linguistically related to these Tocharian speakers, although the “official” language used for administrative purposes was not Tocharian but an MI Prakrit (language of the Niya documents, third century CE), which appears to have been influenced by the phonological structure of Tocharian, sharing with it several features not ordinarily found in an MI Prakrit (Burrow 1935: 667–675):

- (1) absence of voiced stops. The native language of Shan-Shan, Krorainic, lacked the voiced stops *g, j, d, b* so these usually changed to *k, c, t*, and *p* at the beginning of a word and were fricativized (as in G) intervocally (e.g. Kuḡe, Oḡaca, etc., ḡ = [ɣ]). *g, j, d, b* > ḡ, ś, (ḡ), ḡ, *v* (Burrow 1937: §15–16). *d* was sometimes written instead of *t* at the beginning of a word (e.g. *dusya* instead of *tusya*, *dena* instead of *tena*, etc.) as *d* was pronounced as *t*. Intervocally *c* and *j* > -y- or were fricativized (ś, ḡ, i.e. ź, §17).
- (2) absence of aspirated consonants and voiceless glottal fricative *h*. Thus the aspirated stops of Indian words tend to be dropped (*śigra* < S *śighra*, ‘swift’; *agacati* < S *āgacchati*, ‘he comes’), etc.,
- (3) There are no cerebrals (retroflexes) in Tocharian and their appearance in the Niya documents is rare and probably imported (Burrow 1935: 669).
- (4) There is no *v* in Tocharian, only a *w* and *v* in Krorainic occurs only in Sanskrit loanwords; in the native names it is modified to *v̇ = w* (Burrow 1935: 670). In Sanskrit loan-words *p = v* which is also the case in GDhp (Burrow 1937: §20; Brough 1962: §34).
- (5) Sibilants are weakened with *ś* > ź (written as ḡ) and *s* > z (written as *jh* or *s*). In G, single intervocalic *s* was also liable to voicing (Brough 1962: §13) and Khotanese, a neighbouring Middle Iranian language to the west possessed both a /z/ and /z̄/ as part of their phonological inventory which may have influenced the Krorainic pronunciation. There are about forty Iranian loanwords in the Niya Prakrit, indicating a considerable influence from that quarter (Burrow 1937: vii; Mallory and Mair 2000: 278).

Note the similarities with G which also regularly voices or drops intervocalic consonants and those Prakrits (like Pāli) which haphazardly drop or add aspiration to consonantal stops. (e.g. P in Geiger §40, §62; Amg et al., Pischel §206–14). How much these practices were influenced by bilingual Tocharian speakers is

impossible to say; with the limited data we have it is impossible to establish an evolutionary chronology. Tocharian appears to have arrived in the Tarim basin area, even earlier than Indo-Iranian language speakers, with antecedents in the Afanasievo culture of the Altai and Minusinsk regions (3500–2500 BCE), although several other theories are also tenable (Mallory and Mair 2000: 290–296; Mallory 2010: 50–53). There is evidence for the mixing of Tocharian and Indic languages from the third century BCE, when an Indian colony was established in Khotan (the neighbouring kingdom immediately to the west of Shan-Shan), presumably by one of Aśoka’s progeny (Lamotte 1988 [1958]: 257–259).

8.6 Old Sinhalese Prakrit

As stated earlier (Section 5) MI records of the early Buddhist teachings have been preserved only in P, G and BHS. Another Indic language that is relevant to our study however is Old Sinhalese (OS); although geographically distant from the languages we have been discussing it has nevertheless had an important influence on the language of early Buddhism, because of Aśoka’s son Mahinda’s early translation of the *Tipiṭaka* commentaries into this language in the third century BCE (*Cūlavamsa* 37: 228), and their re-translation into P by Buddhaghosa in the fifth century CE. Presumably the source teachings were also translated into OS very early on, as King Devānampiya Tissa (247–207 BCE) was converted by Mahinda’s preaching of the *Cūlahatthipadopamasutta* (MN 1, 175¹³–184²⁰) and other *suttas* which would have been taught in the King’s native language.²⁹

Table 7: Old Sinhalese consonant inventory (from Karunatillake 2001: §2.2.1).

	Labial	Dental	Retroflex	Palatal	Velar
Stops	p	t	ṭ	c	k
	b	d	ḍ	j	g
Nasals	m	n	ṇ		
Semivowel	v				
Lateral		l		y	
Trill			r		
Spirants			s		h

²⁹ In the *Mahāvamsa* 14, v. 65 we read the following:

Laṅkādiṭṭe so satthukappo akappo
Laṅkādhittihāne dvīsu ṭhānesu thero

Sri Lanka was settled by the Indo-Aryans in the sixth century BCE; the original inhabitants of the island were known as the Veddas who spoke a language of unknown genetic affinity (van Driem 2001: 217–242). The first settlers were led by Vijaya who allegedly came from north-west India, while immediately after Vijaya new immigrants arrived from the north-east of India and later from the Tamil speaking central and south parts of the continent. According to Geiger (1935: xxiii–xxiv), Sinhalese had a mixed character. “The base seems to be a Western Dialect brought to the island by the first Aryan colonists. But this base is overgrown with new elements imported into Ceylon at various times, probably already soon after the colonisation and from different parts of the continental India, chiefly from the East (Kaliṅga).” OS is an Old Prakrit which has several similarities with the language of the Aśokan edicts, but also several major differences which may have led to ambiguities in the transmission (from Geiger 1935: xxiv, attested in various cave transcriptions from the second century BCE onwards):

- (1) de-aspiration of all aspirates: e.g. P *Dhammarakkhita* (‘protected by the Dhamma’), OS *Damarakita*; sometimes the aspirated consonant is resolved by splitting, e.g. P *ghāṇa* (‘nose’) > OS *gahaṇa* (Geiger 1938: §36.2). Where the aspirate is retained it does not reflect OI or MI phonology, suggesting that these words, like the other de-aspirates, were pronounced without the aspirate, the phenomenon being due to the “pedantry of scribes” (Paranavithana 1970: xxxi). The de-aspiration feature is probably due to Dravidian influence, because of the proximity of the island to south-east India and Dravidian immigration from that locale.
- (2) the change of *s* > *h*, P *sāṭikā* (‘cloak, mantle’) > OS *hāṭika*; P *posatha* (‘recitation of the Vinaya rules’) > OS *pohata*.
- (3) Shortening of long vowels: P *vāpi* (‘tank’) > OS *vapi* or *vavi*. Whether this was only graphical, as in the omission of long vowels in early continental Brāhmi and Karoṣṭhī script, is not clear (see Paranavithana 1970: xxviii, who maintains that long vowels were not pronounced in actual speech). Other graphical peculiarities are omission of nasals and *anusvāra* before a consonant and replacement of geminate consonants by a single consonant.
- (4) nominative singular in *-e*, as is also found in AMg, Māgadhī and G.

*Dhammaṃ bhāsivā dīpabhāsāya evaṃ
Saddhammotāraṃ kārayi dīpadīpo ti.*

translated by Geiger (1964: 96) as “When thus in the isle of Lanka the peerless *thera* [elder], like unto the Master in the protection of Lanka, had preached the true doctrine in two places, in the speech of the island, he, the light of the island, thus brought to pass the descent of the true faith.”

- (5) single intervocalic stops between vowels are non-phonemic with both lenition and fortition occurring apparently randomly (Paranavithana 1970: xxx); between vowels, stops gradually disappear and by the second to fourth century CE they are replaced by a *y* glide; where they do occur they are usually remnants of consonant clusters (e.g. *k* < *ṛk*, *ṛkh*, *kk*, *kkh*, etc) virtually all of which are assimilated except for nasal + stop, as in #3 above (Karunatillake 2001: §2.1.3 and §2.1.9.1.1b).
- (6) the assimilation of all consonant clusters whether at the beginning of a word or internally.

Since the *buddhadhamma* ('teachings of the Buddha') was translated both into OS (in the third century BCE) and from OS into P (by Buddhaghosa who translated the *aṭṭhakathā* ('commentary') from OS back into P in the fifth century CE), it is not surprising to find "Sinhaleisms" in the Pāli canon, like P *dūta* ('gambling') instead of expected *jūta* < S *dyūta* or *jighacchā* ('hunger') instead of *dighacchā*, because of the medieval Sinhala change of *j* > *d* (von Hinüber 2001: §248). Also in Sinhalese *m* is sometimes substituted for *v* which may explain a word like P *sāmi* ('porcupine') as a substitute for *sāvi* (Geiger 1938: §62.2; von Hinüber 2001: §209) or *Veśamaṇa* < S *Vaiśravaṇa* (Paranavithana 1970: 2); however the alternation of *m* and *v* is also a feature of G and other Prakrits (Brough 1962: §36; Pischel §251, 261). How many of the unetymological de-aspirations of P words (Geiger §40.2, 60.2) are due to Sinhalese influence is unknown; apparently in OS the writing of aspirate vs. non-aspirate consonants was orthographically in free variation (Karunatillake 2001: §2.1.1; e.g. OS *jhaya* written for OI *jāyā*, 'wife' and OS *rajha* written for OI *rājā*, 'king'), which might also account for some of the unetymological aspiration of non-aspirated stops found in the P canon (Geiger §40.1).

9 Discussion

Table 8 compares selected features of Old Indo-Aryan (OIA = OI) and Middle Indo-Aryan (MIA = MI) with other languages which made up the linguistic fabric of India at the time of the Buddha and earlier. Note that the major changes which characterize the diachronic evolution of OIA > MIA and the koinéization of the MI dialects are mirrored in most of these other languages, to wit:

- (1) While voiceless vs. voiced stops were still phonemic in MIA, they were nevertheless weakened (voiceless > voiced) and/or eliminated (voiced > glide or voiced > Ø) in different MI dialects. This process was accelerated by the fact that in PD, Tib, Krorainic, Tocharian and OS, the contrast was non-phonemic, as it was in PM in some environments. The MI *koiné* also eliminated this contrast.

Table 8: Comparison.

Feature	Vedic (OIA)	MIA	MIA <i>koine</i>	Aramaic	Greek <i>koine</i>	Proto-Dravidian	Proto-Munda	Tibetan <i>koine</i>	Krorāṇic	Tocharian	Sinhalese
Voiceless vs. voiced intervocalic stops	Phonemic	Phonemic	Most contrasts eliminated	Phonemic	Phonemic	Non-phonemic	Phonemic in places	Non-phonemic	Non-phonemic	Non-phonemic	Non-phonemic
Weakening of intervocalic stops	Occasionally	Sporadic	Consistently	Spirants become allophones of the dentals	Voiced stops become fricatives	Voiced stops become fricatives	Contrast is neutralized at morpheme boundaries and before vowels	Before nasals unvoiced stops are voiced	Unvoiced stops are voiced or fricativized.	Unvoiced stops are voiced	Unvoiced stops are replaced by a glide
Aspirated stops	Phonemic	Phonemic, replaced with aspirate only in G and other dialects (AMg, Mg, etc.) or fricativized	Consistently replaced by aspirate only	Non-phonemic	Aspirated stops change to fricatives	Non-phonemic, except in borrowed words	Non-phonemic, except in borrowed words	Non-phonemic in Proto and Old Tib.	Non-phonemic	Non-phonemic	Non-phonemic
Monophthongization of Diphthongs	Diphthongs	Monophthongization of diphthongs	Consistently	Monophthongization of diphthongs	Monophthongization of diphthongs	No diphthongs	No diphthongs	No diphthongs	No diphthongs	No diphthongs although Proto Toch. had some	No diphthongs
Phonemic vowel length	Phonemic	Phonemic (but often not written)	Yes	Yes	Non-phonemic	Phonemic	Non-phonemic	Non-Phonemic (Lhasa dialect)	Non-phonemic	Non-phonemic	Unclear

(continued)

Table 8: (continued)

Feature	Vedic (OIA)	MIA	MIA koine	Aramaic	Greek koine	Proto-Dravidian	Proto-Munda	Tibetan koine	Kroranic	Tocharian	Sinhalese
Conjunct consonants (two plosives)	Present	Assimilated sporadically or resolved	Assimilated consistently or resolved	Sporadic assimilation and resolution	Present	Rare internally	Rare internally	Rare internally	Common	Common	Consistently assimilated
Sibilants	Dental, retroflex and palatal	Reduced to dental except in G and Mg	Reduced to single sibilant	Pharyngealized and dental	One dental sibilant	No sibilants	One postalveolar sibilant	Dental and palatal sibilant	Dental retroflex and palatal voiced and unvoiced	Dental, alveolar and palatal	Merged to dental or palatal s
v sound	v; alternates with m in some contexts	v; alternates with m in nasalized contexts	v; alternates with m	No v sound	v	v; alternates with m	No v sound	No v sound	No v sound	No v sound	v; alternates with m

- (2) Sometimes (e.g. in G) the intervocalic stops were further weakened by fricativization, a process probably influenced by Aramaic where interdental spirants were allophones of the dentals and Greek *koine* where voiced stops also became fricatives. This phenomenon also occurred in PD and Krorainic and is found in the MI *koine*.
- (3) In many dialects of MIA, – and in the MI *koine* – the aspirated stop is replaced by an aspirate only. This is a normal change of OIA > MIA, but also one that was probably precipitated and accelerated by the fact that native bilingual speakers of PD, PM, Tib, Krorainic, Tocharian and Sinhalese did not have aspirated stops as part of their phonemic inventory. Aspirated stops were also not phonemic in Aramaic and in Greek *koine* they were often replaced by fricatives (as in G).
- (4) All diphthongs are monophthongized in MIA, and virtually all other Indic and non-Indic languages reinforced this tendency to simplify OIA complex vowels. Similarly, although vowel length was phonemic in MIA and the MI *koine*, its importance was minimized as it was not phonemic in Greek *koine*, PM, Tib, Tocharian or Krorainic, nor was it noted in the early script.
- (5) One of the cardinal features of the evolution of OIA > MIA was the resolution or assimilation of consonant clusters. This process was certainly accelerated by a similar tendency in all the pre-existent non IA languages, viz., PD, PM and the Tib *koine*, where internal conjuncts were rare. Conjuncts would only occur at morpheme boundaries.
- (6) The merger of sibilants in most MIA dialects (except for G) to a single dental *s* (most Prakrits) or palatal *ś* (as in Māgadhī) is another feature of OIA evolution. PD may have influenced this process as sibilants here are non-phonemic, making it difficult for a bilingual Dravidian speaker to hear a sibilant at all.
- (7) The interchange of *-v-* and *-m-* is a process that has been on going since Vedic times, e.g. *-vant* and *-mant*, ‘possessing’; *vhmal* ‘walk crookedly’, *vhval*, *idem*; *vmand* ‘praise’, *vvand* *idem*; *govinda*, *gominda*, proper name; *vśvañc*, *vśmañc* (‘open up’, Wackernagel 2005 [1896]: §177; Bloomfield and Edgerton 1932: §226–§240). This phenomenon was probably also influenced by a similar alternation in Dravidian and the lack of a *-v-* phoneme in Munda, Tib, Tocharian and Chinese.

9.1 Conclusion

A *koine* is a new dialect which reduces linguistic variability by the simplification of marked speech forms. In the case of the MIA *koine* the principal characteristics of the common language were:

- (1) a reduction of the distinction between voiced and voiceless intervocalic stops along with the weakening of intervocalic stops through glide substitution or fricativization, or their elimination altogether.
- (2) the elimination of aspirated stops and their replacement by an aspirate only.
- (3) the elimination of diphthongs and their replacement with a simple vowel.
- (4) the resolution and/or assimilation of consonant clusters.
- (5) the merger of sibilants.
- (6) The interchange of labials *v* and *m*; liquids *l* and *r*; glides *y* and *v*; and the palatals *y* and *j*. Some of this is no doubt due to the levelling of MI dialect differences. Other factors are the influence of non IA languages, like the changeability of the *v* sound, for example, which may be in part due to its non-phonemic status in many of the languages under discussion.

The formation of a *koine* is a very complex process involving a variety of factors: normal OI > MI phonological changes; harmonization of dialectal differences; external pressure from other language contact; and issues of class and social structure, to name some of the principal forces. What has not been very thoroughly examined previously is the impact of other languages – both IA and non-IA – on the development of the MIA inter-language. For the Aryans did not exist in a linguistic void and arguably, in the early centuries of the first millennium BCE they were a minority, outnumbered by the Dravidian, Munda and Tib speaking populations, amongst others. As these bilingual groups struggled to communicate with the increasingly hegemonic MIA speakers, their phonetic constraints accelerated the development of MIA in certain directions, that is, towards harmonization with their own phonological system which lacked such things as phonemic intervocalic voiced stops, aspirated stops, consonant clusters, etc., and had other features (like retroflexes) which IA lacked. The MI *koine* not only facilitated inter-dialect communication, but also inter-language communication by levelling out differences between IA and non-IA phonologies. Whether the Buddha taught in this *koine* or inter-language is impossible to say; however the evidence suggests that his teachings were translated into it at an early time, and from this *koine* into the surviving Prakrits, with various ambiguities resulting due to dialect levelling and simplification, catalyzed and precipitated by diffusional influences from other coeval languages.

Abbreviations

AMg	Ardhamāgadhī
BHS	Buddhist Hybrid Sanskrit
DN	<i>Dīgha Nikāya</i>

G	Gāndhārī
GDhp	<i>Gāndhārī Dhammapada</i> (Brough 1962)
IA	Indo-Aryan
IE	Indo-European
Mhv	<i>Mahāvastu</i> , Senart (1882–97)
MI	Middle Indic
MIA	Middle Indo-Aryan
OED	Oxford English Dictionary
OI	Old Indic
OIA	Old Indo-Aryan
OS	Old Sinhalese
P	Pāli
PD	Proto-Dravidian
PM	Proto-Munda
S	Sanskrit
Sn	<i>Sutta Nipāta</i> , Andersen and Smith 2010 [1913]
Tib	Tibetan

‘(apostrophe) = intervocalic > Ø; in G = *alif* [the letter a] as a syllable divider per Brough (1962: §37).

References

- Alsdorf, Ludwig. 1980. Ardha-Māgadhī. In Heinz Bechert (ed.), *Die Sprache der ältesten buddhistischen Überlieferung. Abhandlungen der Akademie der Wissenschaften in Göttingen, Philologisch-Historische Klasse, Dritte Folge, Nr. 117.* 17–23.
- Ananthanarayana, H. S. 1991. Assimilation and Sandhi in MIA: A case of convergence between Indo-Aryan and Dravidian. In B. Lakshmi Bai & B. Ramakrishna Reddy, (eds.), *Studies in Dravidian and general linguistics, A festschrift for Bh. Krishnamurti*, 256–262. Hyderabad: Centre of Advanced Study in Linguistics, Osmania University.
- Andersen, Dines & Helmer Smith. 2010 [1913]. *Sutta Nipata*. Oxford: Pali Text Society
- Andronov, M. 1968. *Dravidian and Aryan: From the typological similarity to the similarity of forms. Two lectures on the historicity of language families*. Annamalainagar: Annamalai University.
- Bailey, H. W. 1946. Gāndhārī. *Bulletin of the School of Oriental (and African) Studies* 11. 1943–46. 764–797.
- Bloomfield, Maurice & Franklin Edgerton. 1932. *Vedic Variants, Volume II Phonetics*. Philadelphia: Linguistic Society of America.
- Bodhi, B. 2007 [1978]. *The all-embracing net of views: The Brahmajāla Sutta and its commentaries*. Kandy: Buddhist Publication Society.
- Bonnerjea, Biren. 1936. Phonology of some Tibeto-Burman dialects of the Himalayan region. *T'oung Pao Second Series* 32(4). 238–258.
- Brough, John. 1962. *The Gāndhārī Dharmapada*. London: Oxford University Press.
- Bryant, Edwin. 2001. *The quest for the origins of Vedic culture: The Indo-Aryan migration debate*. New York: Oxford University Press.
- Burrow, Thomas. 1935. Tocharian elements in Kharoṣṭhi documents. *Journal of the Royal Asiatic Society* 4 (4). 667–675.

- Burrow, Thomas. 1937. *The language of the Kharoṣṭhi documents from Chinese Turkestan*. Cambridge: Cambridge University Press.
- Caillat, C. 1968. Isipatana Migadāya. *Journal Asiatique* 256. 177–183.
- Campbell, L. 2004. *Historical linguistics*. Cambridge: The MIT Press.
- Cardona, G. 1985. Indo-Iranian languages. In P. W. Goetz. (ed.) *The new Encyclopaedia Britannica*, 612–624. Chicago: Encyclopaedia Britannica Inc.
- Childers, R. C. 1875. *A dictionary of the Pali language*. London: Trubner & Co.
- Colvin, Stephen. 2007. *Historical Greek reader: Mycenaean to the Koiné*. Oxford: Oxford University Press.
- DeLancey, Scott. 2003a. Classical Tibetan. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 255–269. London & New York: Routledge.
- DeLancey, Scott. 2003b. Lhasa Tibetan. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 270–288. London & New York: Routledge.
- Denwood, Philip. 1999. *Tibetan*. Amsterdam & Philadelphia: John Benjamins.
- Deshpande, M. M. 1979. Genesis of Ṛgvedic retroflexion: A historical and sociolinguistic investigation. In M. M. Deshpande & P. E. Hook (eds.), *Aryan and non-Aryan in India*, 235–315. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.
- Deshapande, M. M., 1993. *Sanskrit & Prakrit sociolinguistic issues*. Delhi: Motilal Banarsidass Publishers.
- Driem, George van. 2001. *Languages of the Himalayas: An ethnolinguistic handbook of the Greater Himalayan region*. Leiden, Boston & Köln: Brill.
- Edgerton, Franklin. 1936. The Prakrit underlying Buddhistic hybrid Sanskrit. *Bulletin of the School of Oriental Studies*, University of London. 8(2/3). 501–516.
- Edgerton, Franklin. 1988 [1953]. *Buddhist hybrid Sanskrit grammar and dictionary*. Delhi: Motilal Banarsidass Publishers.
- Emeneau, Murray B. 1956. India as a linguistic area. *Language* 32. 3–16.
- Emeneau, Murray B. 1974. The Indian linguistic area revisited. *International Journal of Dravidian Linguistics* 3. 92–134.
- Emeneau, Murray B. 1978. Linguistic area: Introduction and continuation. *Language* 54. 201–210.
- Geiger, Wilhelm. 1916. *Pāli Literatur und Sprache*. Strassburg: Verlag from Karl. J. Trübner.
- Geiger, Wilhelm. 1935. *A dictionary of the Sinhalese language*. Colombo: The Royal Asiatic Society, Ceylon Branch.
- Geiger, Wilhelm. 1938. *A grammar of the Sinhalese language*. Londonolombo: The Royal Asiatic Society, Ceylon Branch.
- Geiger, Wilhelm. 1964. *The Mahavamsa or the great chronicle of Ceylon*. London: London Pali Text Society.
- Geiger, Wilhelm. 2005 [1916]. *A Pāli grammar, translated into English by Batakriṣṇa Bhosh*, revised and edited by K. R. Norman. Oxford: The Pali Text Society.
- Ghosh, A. 2008. Santali. In G. D. S. Anderson (ed.), *The Munda languages*. 11–98. New York: Routledge Taylor & Francis Group.
- Hahn, Michael. 2002. *Textbook of classical literary Tibetan, translated and revised by Ulrich Pagel*. London. <http://isites.harvard.edu/icb/icb.do?keyword=k41868&pageid=icb>.
- Hill, Nathan. 2010. An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society* 108(2). 110–125.
- Hinüber, Oskar von. 1983. The oldest literary language of Buddhism. *Saeculum* 34. 1–9. Also published in *Selected papers on Pāli studies*, 177–194. Oxford: The Pali Text Society, 2005.

- Hinüber, Oskar von. 2001. *Das Ältere Mittelindisch im Überblick*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Hock, Hans Henrich. 1996. Pre-Ṛgvedic convergence between Indo-Aryan (Sanskrit) and Dravidian? A survey of the issues and controversies. In Jan E. M. Houben (ed.), *Ideology and status of Sanskrit: Contributions to the history of the Sanskrit language*, 17–58. Leiden: E. J. Brill.
- Horner, Isaline B. 2001 [1938]. *Book of the discipline*, vol. 1, London: Pali Text Society. Six volumes.
- Horrocks, G. 2010. *Greek: A history of the language and its speakers*. Oxford: Wiley-Blackwell.
- Karunatilake, W. S. 2001. *Historical phonology of Sinhala: From old Indo-Aryan to the 14th Century A.D.* Colombo: S. Godage & Brothers.
- Keith, A. B. 1971 [1920]. *Rigveda Brahmanas: The Aitareya and Kauṣītaki Brāhmaṇas of the Rigveda*. Delhi: Motilal Banarsidass.
- Kerswill, Paul & Anne Williams. 2000. Creating a new town Koine: Children and language change in Milton Keynes. *Language in Society* 29. 65–115.
- Krause, Todd B. & Jonathan Slocum. 2007–2010. Tocharian Online. Linguistic Research Center. University of Texas at Austin. <http://www.utexas.edu/cola/centers/lrc/eieol/tokol-1-X.html#L211> (accessed December 2014).
- Kuiper, Franciscus B. J. 1965. Consonant variation in Munda. *Lingua* 14. 54–87.
- Kuiper, Franciscus B. J. 1967. The genesis of a linguistic area. *Indo-Iranian Journal* 10. 81–102.
- Kulikov, L. 2013. Language vs. grammatical tradition in Ancient India: How real was Pāṇinian Sanskrit. *Folia Linguistica Historica* 34. 59–91.
- Lamotte, Étienne. 1988 [1958]. *History of Indian Buddhism from the origins to the Śāka Era, translated from the French by Sara Webb-Boin*. Louvain-la-Neuve: Université Catholique de Louvain Institut Orientaliste.
- Lévi, S. 1912. Observations sur une langue précanonique du bouddhisme. *Journal Asiatique* 20. 495–514.
- Levman, Bryan G. 2008–2009. *Sakāya niruttiyā* Revisited. *Bulletin des Études Indiennes* 26–27. 33–51.
- Levman, Bryan G. 2010. Aśokan phonology and the language of the earliest Buddhist tradition. *Canadian Journal of Buddhist Studies* 6. 57–88.
- Levman, Bryan G. 2011. What does the Pāli phrase *pahitatta* mean? *Thai International Journal for Buddhist Studies* 3. 57–75.
- Levman, Bryan G. 2012. Lexical ambiguities in the Buddhist teachings: An example & methodology. *International Journal for the Study of Humanistic Buddhism* (2). 35–54.
- Levman, Bryan G. 2013. Cultural remnants of the indigenous peoples in the Buddhist Scriptures. *Buddhist Studies Review* 30(2). 145–189.
- Levman, Bryan G. 2014. *Linguistic ambiguities: The transmissional process, and the earliest recoverable language of Buddhism*. Toronto, University of Toronto. Ph.D Dissertation.
- Lubotsky, A. 1995. Sanskrit *h* < **dh*, *bh*. In N. V. Gurov & Y. V. Vasil'kov (eds.), *Śthāpakaśrāddham: Professor G. A. Zograph Commemorative Volume*, 122–124. St. Petersburg: Peterburgskoe Vostokovedenie
- Lüders, Heinrich. 1954. *Beobachtungen über die Sprache des Buddhistischen Urkanons*. Berlin: Akademie-Verlag.
- Mallory, J. P. 2010. Bronze Age languages of the Tarim Basin. *Expedition* 523. 44–53.
- Mallory, J. P. & Victor H. Mair. 2000. *The Tarim mummies*. London: Thames & Hudson.
- Masica, C. P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman*. Berkeley & Los Angeles: University of California Press.
- Mayrhofer, Manfred. 1963. *Kurzgefasstes etymologisches Wörterbuch des Altindischen: A Concise Etymological Sanskrit Dictionary*. Heidelberg: Carl Winter – Universitätsverlag.
- Mehendale, Madhukar A. 1968. *Some aspects of Indo-Aryan linguistics*. Bombay: University of Bombay.
- Mufwene, Salikoko S. 1985. Koine. In P. W. Goetz (ed.), *Encyclopaedia Britannica*. Chicago: Encyclopaedia Britannica, Inc. <http://www.britannica.com/EBchecked/topic/321152/koine> (accessed December 2014).
- Mylius, Klaus. 2003. *Wörterbuch Ardhamāgadhī-Deutsch*. Wichtrach: Institut für Indologie.
- Norman, Kenneth Roy. 1980a. Four etymologies from the Sabhiya-sutta. In Somaratna Balasooriya et al. (eds.), *Buddhist studies in honour of Walpola Rahula, 173–184*. London: Gordon Fraser. Also published in *Collected Papers 2* (Oxford: Pali Text Society, 1991), 148–161.
- Norman, Kenneth Roy. 1980b. The dialects in which the Buddha preached. In Heinz Bechert (ed.), *Die Sprache der ältesten buddhistischen Überlieferung. Abhandlungen der Akademie der Wissenschaften in Göttingen, Philologisch-Historische Klasse, Dritte Folge, Nr. 117*. 61–77. Reprinted in *Collected Papers 2*, 128–147. Oxford: Pali Text Society.
- Norman, K. R. 1983. *Pāli literature*. Wiesbaden: Otto Harrassowitz.
- Norman, Kenneth Roy. 1989. Dialect Forms in Pali. In C. Caillat, (ed.) *Dialectes dans les littératures indo-aryennes*, 369–392. Paris: Institut de Divilisation Indienne. Also available in *Collected Papers 4* (Oxford: The Pali Text Society, 1993), 46–71.
- Norman, Kenneth Roy. 1993. The languages of early Buddhism. In *Premier Colloque Étienne Lamotte* (Bruxelles et Liège 24–27 septembre 1989), 83–99. Louvain-la-neuve: Université Catholique de Louvain Institut Orientaliste. Also available in *Collected Papers 5* (Oxford: The Pali Text Society, 1994), 146–168.
- Norman, Kenneth Roy. 2006. *A philological approach to Buddhism: The Bukkyō Dendō Kyōkai Lectures 1994*. Lancaster: The Pali Text Society
- Norman, Kenneth Roy. 2012. The languages of the composition and transmission of the Aśokan inscriptions. In P. Olivelle, J. Leoshko & H. P. Ray (eds.), *Reimagining Aśoka: Memory and history*, 38–62. New Delhi: Oxford University Press.
- Oldenberg, H. 1882. *Buddha: His life, his doctrine, his order, translated from the German by William Hoey*. London: Williams and Norgate.
- Osada, Toshiki. 2008. Mundari. In Gregory D. S. Anderson (ed.), *The Munda languages*, 99–164. New York: Routledge Taylor & Francis Group.
- Paranavithana, Senarath. 1970. *Inscriptions of Ceylon, Volume 1, containing cave inscriptions from 3rd Century B.C. to 1st Century A.C. and other inscriptions in the early Brāhmī script*. Ceylon: Department of Archaeology Ceylon.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer Historischen Laulehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
- Pischel, Richard. 1981 [1900]. *Comparative grammar of the Prākṛit languages*, translated from the German by Subhadra Jhā. Delhi: Motilal Banarsidass.
- Reich, David, Kumarasamy Thangaraj, Nick Patterson, Alkes L. Price & Lalji Singh. 2009. Reconstructing Indian population history. *Nature* 461. 489–495.
- Rhys Davids, T. W. 1908. *Early Buddhism*. London: Archibald Constable & Co.
- Roth, Gustav. 1980. Particular features of the language of the Ārya-Mahāsāṃghika-Lokottaravādins and their importance for early Buddhist tradition. In Heinz Bechert (ed.),

- Die Sprache der ältesten buddhistischen Überlieferung. Abhandlungen der Akademie der Wissenschaften in Göttingen, Philologisch-Historische Klasse, Dritte Folge, Nr. 117.* 78–135.
- Salomon, R. 1998. *Indian epigraphy: A guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages*. New Delhi: Mushiram Manoharlal Publishers Pvt. Ltd.
- Salomon, R. 2000. *A Gāndhārī version of the Rhinoceros Sūtra*. Seattle & London: University of Washington Press.
- Schayer, S. 1935. Notes and queries on Buddhism. *Rocznik Orientalistyczny* 11. 206–13.
- Segert, Stanislav. 1997. Old Aramaic phonology. In A. S. Kaye (ed.), *Phonologies of Asia and Africa*, vol. 1. 115–125. Winona Lake, IN: Eisenbrauns.
- Senart, E. 1882–1897. *Mahāvastu* 1, 2, and 3. Paris: Imprimerie Nationale.
- Sharma, D. D. 2003. *Munda sub-stratum of Tibeto-Himalayan languages*. New Delhi: Mittal Publications.
- Siegel, Jeff. 1985. Koines and koineization. *Language in Society* 14. 357–378.
- Smith, H. 1952. Le futur moyen indien. *Journal Asiatique* 240. 169–183.
- Southworth, Franklin C. 1974. Contact and convergence in South Asian languages. In Franklin C. Southworth & Mahadev L. Apte (eds.), *International Journal of Dravidian Linguistics* 3(1). 201–223.
- Southworth, Franklin C. 2005. *Linguistic archaeology of South Asia*. London: Routledge.
- Sprigg, Richard Keith. 1972. A polysystemic approach, in Proto-Tibetan reconstruction, to tone and syllable-initial consonant clusters. *Bulletin of the School of Oriental (and African) Studies* 35(3). 546–587.
- Steever, Sanford B. 1998. Introduction to the Dravidian languages. In S. B. Steever (ed.), *The Dravidian languages*, 1–39. London & New York: Routledge.
- Thomas, E. J. 1937. *Tathāgata and Tahāgaya*. *Bulletin of the School of Oriental Studies, London* 8. 781–788.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Basil Blackwell Ltd.
- Wackernagel, Jacob. 2005 [1896]. *Altindische Grammatik, Volume I Lautlehre*. Göttingen: Dandenhoeckund Ruprecht, reprint by Elibron Classic Series.
- Weinreich, U., 1967 [1953]. *Languages in contact*. The Hague: Mouton.
- Witzel, Michael. 1999. Substrate languages in Old Indo-Aryan (Ṛgvedic, Middle and Late Vedic). *Electronic Journal for Vedic Studies* 5. 1–67.
- Witzel, Michael. 2001. Autochthonous Aryans? The evidence from Old Indian and Iranian texts. *Electronic Journal of Vedic Studies* 7(3). 1–117.
- Woolner, Alfred C. 1926–1928. Prakritic and non-Aryan strata in the vocabulary of Sanskrit. In *Sir Asutosh Memorial Volume*, 65–71. Patna: J. N. Samaddar.
- Wüst, Walter. 1957. Ṭhakkura-, m. zur Problematik der Indoarischen Zerebralisation und des Lehn-sprachen-Einflusses. *Mitteilungen zur idg., vornehmlich indo-iranischen Wortkunde sowie zur holothetischen Sprachtheorie* 3. 5–98.
- Zide, Norman. 1969. Munda and non-Munda austroasiatic languages. In T. A. Sebeok (ed.), *Current trends in linguistics*, 411–430. The Hague, Paris: Mouton.
- Zvelebil, Kamil V. 1990. *Dravidian linguistics: An introduction*. Pondicherry: Pondicherry Institute of Linguistics and Culture.